PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 : C12Q 1/68

A1

(11) International Publication Number: WO 93/00446

(43) International Publication Date: 7 January 1993 (07.01.93)

(21) International Application Number: PCT/US92/05476

(22) International Filing Date: 26 June 1992 (26.06.92)

(30) Priority data:
723,618 27 June 1991 (27.06.91) US

(60) Parent Application or Grant
(63) Related by Continuation
US 723,618 (CIP)
Filed on 27 June 1991 (27.06.91)

(71) Applicant *(for all designated States except US):* GENELABS INCORPORATED [US/US]; 505 Penobscot Drive, Redwood City, CA 94063 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only) :* EDWARDS, Cynthia, A. [US/US]; 2021 Oakley Avenue, Menlo Park, CA 94025 (US). CANTOR, Charles, R. [US/US]; 640 Panoramic Way, Berkeley, CA 94707 (US). ANDREWS, Beth, M. [US/US]; 24 Franklin Street, Watertown, MA 02172 (US).

(74) Agent: BENSON, Robert, H.; Genelabs Incorporated, 505 Penobscot Drive, Redwood City, CA 94063 (US).

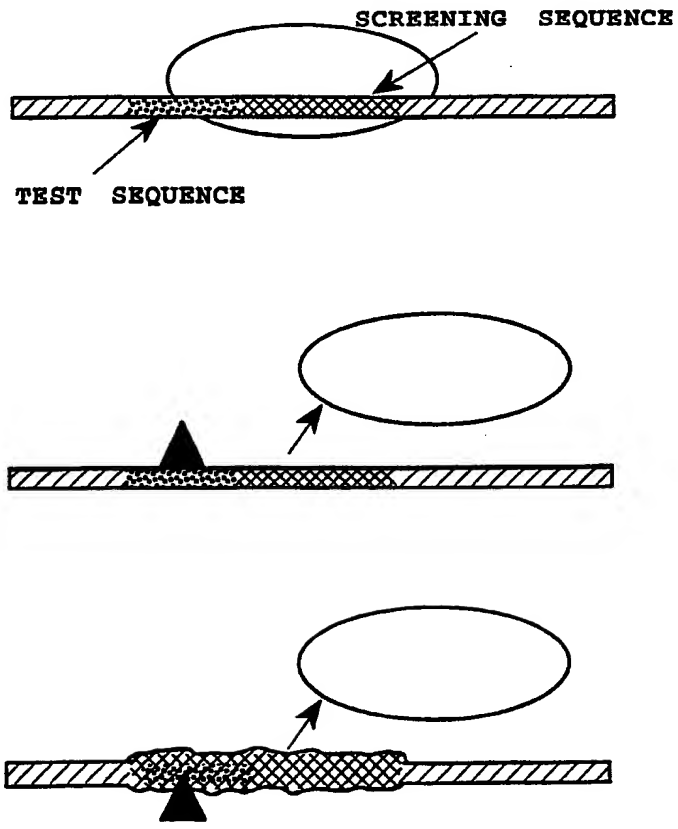
(81) Designated States: AU, CA, JP, KR, US, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LU, MC, NL, SE).

Published
With international search report.

(54) Title: SCREENING ASSAY FOR THE DETECTION OF DNA-BINDING MOLECULES

(57) Abstract

The present invention defines an assay useful for screening libraries of synthetic or biological compounds for their ability to bind specific DNA test sequences. The assay is also useful for determining the sequence specificity and relative DNA-binding affinity of DNA-binding molecules for any particular DNA sequence. The assay is a competition assay in which binding of a test molecule to a DNA test sequence changes the binding characteristics of a DNA-binding protein to its binding sequence. When such a test molecule binds the test sequence the equilibrium of the DNA:protein complexes is disturbed, generating changes in the ratio between unbound DNA and DNA:protein complexes. The assay is versatile in that any test sequence can be tested by placing the test sequence adjacent to a defined protein binding DNA screening sequence.



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	FI	Finland	ML	Mali
AU	Australia	FR	France	MN	Mongolia
BB	Barbados	GA	Gabon	MR	Mauritania
BE	Belgium	GB	United Kingdom	MW	Malawi
BF	Burkina Faso	GN	Guinea	NL	Netherlands
BG	Bulgaria	GR	Greece	NO	Norway
BJ	Benin	HU	Hungary	PL	Poland
BR	Brazil	IE	Ireland	RO	Romania
CA	Canada	IT	Italy	RU	Russian Federation
CF	Central African Republic	JP	Japan	SD	Sudan
CG	Congo	KP	Democratic People's Republic of Korea	SE	Sweden
CH	Switzerland			SN	Senegal
CI	Côte d'Ivoire	KR	Republic of Korea	SU	Soviet Union
CM	Cameroon	LI	Liechtenstein	TD	Chad
CS	Czechoslovakia	LK	Sri Lanka	TG	Togo
DE	Germany	LU	Luxembourg	US	United States of America
DK	Denmark	MC	Monaco		
ES	Spain	MG	Madagascar		

SCREENING ASSAY FOR THE DETECTION OF DNA-BINDING MOLECULES

Field of the Invention

The present invention relates to a method, a system, and a kit useful for the identification of molecules that specifically bind to defined nucleic acid sequences.

References

Ambinder, R.F., *et al.*, *J. Virol.* 65:1466-1478, 1991.

Ausubel, F. M., et al., Current Protocols in Molecular Biology, John Wiley and Sons, Inc., Media PA.

Banerji, S.S., et al., *Mol. Cell Biol.* 11:4074-4087, 1991.

Beal, P.A., *et al.*, *Science* 251:1360-1363, 1991.

Birg, F., *et al.*, *Nucl. Acids Res.* 18:2901-2908, 1990.

Chaiet, L., et al., Arch. Biochem. Biophys. 106:1 (1964).

Chang, H.-K, et al., *Mol. Cell. Biol.* Nov.:5189-5197, 1989.

Chang, H.-K,, et al., Mol. Cell. Biol. November:-5189-5197 (1989).

Chen, K-X., et al., J. Biomol. Struct. Dyn. 3:445-466 (1985).

Chin, M.T., et al., *J. Virol.* 63:2967-2976, 1989.

Cooney, M., *t al.*, *Science* 241:456-459, 1988.

Courtois, G., et al., Proc.Natl.Acad.Sci.USA

85:7937-7941 (1988).

Cullinane, C., *et al.*, *FEBS lett.* 293:195-198, 1991.

Edwards, C.A., *et al.*, *J.Mol.Biol.* 180:73-90, 1984.

Edwards, C.A. and Crabtree, G.R. Molecular regulation of T lymphocyte activation. In: *Advances in Regulation of Cell Growth, Volume I: Regulation of Cell Growth and Activation*, edited by Mond, J.J., Weiss, A. and Cambier, J.C. New York: Raven Press, 1989, p. 91-118.

Elias, P., *et al.*, *J. Biol. Chem.* 265:17167-17173, 1990.

Elias, P., et al., Proc. Natl. Acad. Sci. USA 85:2959-2963 (1988).

Fried, M.G., et al., Nuc. Acid. Res. 9:6505 (1981).

Froehler, et al., Nucleic Acids Res. 16:4831 (1988).

Galas, D., et al., Nuc.Acid Res. 5:3157-3170 (1981).

Garner, M.M., et al., Nuc. Acid. Res. 9:3047 (1981).

Gaugain, B., et al., *Biochemistry* 17:5078, 1978.

Gaugain, B., *et al.*, *Biochemistry* 17:5071, 1978.

Gessner, R.V., et al., Biochemistry 24:237-240 (1985).

Gilbert, D.F., et al., Proc. Natl. Acad. Sci. USA 86:3006 (1988).

Gilman, A. G., et al., eds., The Pharmacological Basis of Therapeutics, Eighth Edition, Pergamon Press (1990).

Goldin, A.L., et al., J. Virol. 38:5-58 (1981).

Green, N.M., Adv. Protein Chem. 29:85 (1975).

Greene, W.C. Regulation of HIV-1 gene expression. *Annu. Rev. Immunol.* 8:453-475, 1990.

Griffin, L.C., et al., *Science* 245:967-971, 1989.

Gross, D.S., *et al.*, *Annu.Rev.Biochem.* 57:159-197, 1988.

Harlow, E., et al., Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press (1988).

Helene, C., et al., *Genome* 31:413-420, 1989.

Helene, C., *t al.*, *Biochim. Biophys. Acta* 1049:99-125, 1990.

Jain, S.C., et al., J. Mol. Biol. 68:1-20 (1972).

Kadonaga, J.T., PNAS 83:5889-5893 (1986).

Koff, A., et al., J. Virol. 62:4096-4103 (1988).

Kuhlmann, K.F., *et al.*, *Nucl. Acids Res.* 5:2629, 1978.

Laugaa, P., *et al.*, *Biochemistry* 23:1336, 1985.

Le Pecq, J.B., et al., *Proc. Natl. Acad. Sci. U.S.A.* 72:2915-2919, 1975.

Luck, G., et al., Nucl. Acids Res. 1:503 (1974).

Luckow, V.A., et al., Virology 170:31 (1989).

Maher III, L.J., et al., *Science* 245:725-730, 1989.

Maniatis, T., et al. Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory (1982).

McGeoch, D. J., et al., J. Virology 62(2):444 (1988).

McGeoch, D.J., et al., *J. Virol.* 62:444-453, 1988.

Miller, P.S., et al., U.S. Patent No. 4,507,433, issued 26 March 1985.

Miller, et al., Biochemistry 18:5134 (1979).

Miller, et al., J Biol Chem 255:6959 (1980).

Miller, et al., Biochimie 67:769 (1985).

Nielsen, P., *et al.*, Science 254:1497-1500 (1991).

Olivo, P.D., et al., Proc. Natl. Acad. Sci. USA 85:5414-5418 (1988).

Olivo, P.D., et al., J. Virology 3:196-204 (1989).

Pelaprat, D., *et al.*, *J. Med. Chem.* 23:1336-1343, 1980.

Perouault, L., et al., *Nature* 344:358-360, 1990.

Phillips, D.R., *et al.*, *Anti-Cancer Drug Design* 5:21-29, 1990.

Phillips, D.R., *Biochemistry* 25:7355-7362, 1986.

Polinksy, B., et al., PNAS 72:3310-4 (1975).

Quigley, G.J., et al., Science 232:1255-1258 (1986).

Reisman, D., *et al.*, *Mol. Cell. Biol.* 5:1822-1832, 1985.

Salas, X., t al., *FEBS lett.* 292:223-228, 1991.

Sambrook, J., et al., In Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Vol. 2

(1989).

Schmidt, A., *et al.*, *J. Virol.* 64:4037-4041, 1990.

Sherman, S.E., et al., Chem. Rev. 87:1153 (1987).

Siebenlist, U., et al., Proc. Natl. Acad. Sci. USA 77:122-126 (1980).

Skorobogaty, A.,, et al., *Anti-Cancer Drug Design* 3:41-56, 1988.

Smith, D.B., et al., Gene 67:31 (1988).

Sobell, H.M., et al., J. Mol. Biol. 68:21-34 (1972).

Sobell, H.M., Prof. Nucl. Acid. Res. Mol. Biol. 13:153-190 (1973).

Stow, N.D., *et al.*, *J. Gen. Virol.* 67:1613-1623, 1986.

Stow, N.D., et al., Virology 130:427-438 (1983).

Strobel, S.A., *et al.*, *Science* 249:73-75, 1990.

Summers, M.D., et al., A manual of Methods for Baculovirus Vectors and Insect Cell Culture Procedures, Texas Agricultural Experimental Station Bulletin, No. 1555 (1987).

Summerton, J.E., *et al.*, U.S. Patent No. 5,034,506, issued 23 July 1991.

Summerton, J.E., *et al.*, P.C.T. International Application PCT/US86/00544, International Publication No. WO 86/05518, published 25 September 1986.

Tullius, T.D., Ann. Rev. Biophys. Biochem. 18:213-237 (1989).

Tullius, T.D., *Annu. Rev. Biophys. Biophys. Chem.* 18:213-237, 1989.

Wartel, R.M., et al., J. Biol. Chem. 15:285-318 (1975).

Weir, H.M., et al., Nucl. Acids Res. 17:1409-1425 (1989).

Woodbury, C.P., et al., Biochemistry 22(20):4730-4737 (1983).

Wu, C.A., et al., J. Vir 1. 62:435-443 (1988).

Young, S.L., et al., *Proc. Natl. Acad. Sci. U.S.A.*

88:10023-10026, 1991.

Zein, N., et al., Science 240:1198 (1988).

Zimmer, C., Pros. Nucl. Acid Res. Mol. Biol. 15:285-318 (1975).

Background of the Invention

Several classes of small molecules that interact with double-stranded DNA have been identified. Many of these small molecules have profound biological effects. For example, many aminoacridines and polycyclic hydrocarbons bind DNA and are mutagenic, teratogenic, or carcinogenic. Other small molecules that bind DNA include: biological metabolites, some of which have applications as antibiotics and antitumor agents including actinomycin D, echinomycin, distamycin, and calicheamicin; planar dyes, such as ethidium and acridine orange; and molecules that contain heavy metals, such as cisplatin, a potent antitumor drug.

Most known DNA-binding molecules do not have a known sequence binding preference. However, there are a few small DNA-binding molecules that preferentially recognize specific nucleotide sequences, for example: echinomycin preferentially binds the sequence [(A/T)CGT]/[ACG(A/T)] (Gilbert et al.); cisplatin covalently cross-links a platinum molecule between the N7 atoms of two adjacent deoxyguanosines (Sherman et al.); and calicheamicin preferentially binds and cleaves the sequence TCCT/AGGA (Zein et al.).

The biological response elicited by most therapeutic DNA-binding molecules is toxicity, specific only in that these molecules may preferentially affect cells that are more actively replicating or transcribing DNA than other cells. Targeting specific sites may significantly decrease toxicity simply by reducing the number f potential binding sites in the DNA. As specificity for l nger sequences is acquired, the nonsp cific toxic ffects due to DNA-binding

may decrease. Many therapeutic DNA-binding molecules initially identified based on their therapeutic activity in a biological screen have been later determined to bind DNA.

Therefore, there is a need for an *in vitro* assay useful to screen for DNA-binding molecules. There is also a need for an assay that allows the discrimination of sequence binding preferences of such molecules. Additionally, there is a need for an assay that allows the determination of the relative affinities of a DNA-binding molecule for different DNA sequences. Finally, there is a need for therapeutic molecules that bind to specific DNA sequences.

Summary of the Invention

The present invention provides a method for screening molecules or compounds capable of binding to a selected test sequence in a duplex DNA. The method involves adding a molecule to be screened, or a mixture containing the molecule, to a test system. The test system includes a DNA binding protein that is effective to bind to a screening sequence, i.e. the DNA binding protein's cognate binding site, in a duplex DNA with a binding affinity that is preferably substantially independent of the sequences adjacent the binding sequence -- these adjacent sequences are referred to as test sequences. But, the DNA binding protein is sensitive to binding of molecules to such test sequence, when the test sequence is adjacent the screening sequence. The test system further includes a duplex DNA having the screening and test sequences adjacent one another. Also, the binding protein is present in an amount that saturates the screening sequence in the duplex DNA. The test molecule is incubated in contact with the test system for a period sufficient to permit binding of the molecule being t sted t the test sequence in the duplex DNA. The amount of binding protein bound to the duplex DNA

is compared before and after the addition of the test molecule or mixture.

Candidates for the screening sequence/binding protein may be selected from the following group: EBV origin of replication/EBNA, HSV origin of replication/UL9, VZV origin of replication/UL9-like, HPV origin of replication/E2, interleukin 2 enhancer/NFAT-1, HIV-LTR/NFAT-1, HIV-LTR/NFkB, HBV enhancer/HNF-1, fibrinogen promoter/HNF-1, lambda o_L-o_R/cro, and essentially any other DNA:protein interactions.

A preferred embodiment of the present invention utilizes the UL9 protein, or DNA-binding proteins derived therefrom, and its cognate binding sequence SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:17, or SEQ ID NO:15.

The test sequences can be any combination of sequences of interest. The sequences may be randomly generated for shot-gun approach screening or specific sequences may be chosen. Some specific sequences of medical interest include the following sequences involved in DNA:protein interactions: EBV origin of replication, HSV origin of replication, VZV origin of replication, HPV origin of replication, interleukin 2 enhancer, HIV-LTR, HBV enhancer, and fibrinogen promoter. Furthermore, a set of assay test sequences comprised of all possible sequences of a given length could be tested (eg., all four base pair sequences).

In the above method, comparison of protein-bound to free DNA can be accomplished using any detection assay, preferably, a gel band-shift assay, a filter-binding assay, or a capture/detection assay.

In one embodiment of the DNA capture/detection assay, in which the DNA that is not bound to protein is captured, the capture system involves the biotinylation of a nucleotid within the screening sequence (i) that does not eliminat the protein's ability to bind to the screening

sequence, (ii) that is capable of binding streptavidin, and (iii) where the biotin moiety is protected from interactions with streptavidin when the protein is bound to the screening sequence. The capture/detection assay also involves the detection of the captured DNA.

In another embodiment of the DNA capture/detection assay, the capture system in which the DNA:protein complexes are captured, the capture system involves the use of nitrocellulose filters under low salt conditions to capture the protein-bound DNA while allowing the non-protein-bound DNA to pass through the filter.

The present invention also includes a screening system for identifying molecules that are capable of binding to a test sequence in a duplex DNA sequence. The system includes a DNA binding protein that is effective to bind to a screening sequence in a duplex DNA with a binding affinity that is substantially independent of a test sequence adjacent the screening sequence. The binding of the DNA protein is, however, sensitive to binding of molecules to the test sequence when the test sequence is adjacent the screening sequence. The system includes a duplex DNA having the screening and test sequences adjacent one another. Typically, the binding protein is present in an amount that saturates the screening sequence in the duplex DNA. The system also includes means for detecting the amount of binding protein bound to the DNA.

As described above the test sequences can be any number of sequences of interest.

The screening sequence/binding protein can be selected from known DNA:protein interactions using the criteria and guidance of the present disclosure. It can also be applied to DNA:Protein interactions later discovered.

A preferred embodiment of the screening system of the present invention includes the UL9 protein, or DNA-binding protein derived therefrom (e.g., the truncated UL9 protein

designated UL9-COOH). In this embodiment the duplex DNA has (i) a screening sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:17 and SEQ ID NO:15, and (ii) a test sequence adjacent the screening sequence, where UL9 is present in an amount that saturates the screening sequence. The system further includes means for detecting the amount of UL9 bound to the DNA, including, band-shift assays, filter-binding assays, and capture/detection assays.

The present disclosure describes the procedures needed to test DNA:protein interactions for their suitability for use in the screening assay of the present invention.

The present invention further defines DNA capture systems and detection systems. Several methods are described. A filter binding assay can be used to capture the DNA:protein complexes or, alternatively, the DNA not bound by protein can be captured by the following method. In the first part of this system, the cognate DNA binding site of the DNA binding protein is modified with a detection moiety, such as biotin or digoxigenin. The modification must be made to the site in such a manner that (i) it does not eliminate the protein's ability to bind to the cognate binding sequence, (ii) the moiety is accessible to the capturing agent (e.g., in the case of biotin the agent is streptavidin) in DNA that is not bound to protein, and (iii) where the moiety is protected from interactions with the capture agent when the protein is bound to the screening sequence.

In the second part of this system, the target oligonucleotide is labelled to allow detection. Labelling of the target oligonucleotide can be accomplished by standard techniques such as radiolabelling. Alternatively, a moiety such as digoxigenin can be incorporated in the target oligonucleotide and this moiety can then be detected after capture.

Three embodiments of the capture/detection system described by the present disclosure are as follows:

(i) the target oligonucleotide (containing, for example, the screening and test sequences) -- modification of the cognate binding site with biotin and incorporation of digoxigenin or radioactivity (eg., ^{35}S or ^{32}P); capture of the target oligonucleotide using streptavidin attached to a solid support; and detection of the target oligonucleotide using a tagged anti-digoxigenin antibody or radioactivity measurement (eg., autoradiography, counting in scintillation fluor, or using a phosphoimager).

(ii) the target oligonucleotide -- modification of the cognate binding site with digoxigenin and incorporation of biotin or radioactivity; capture of the target oligonucleotide using an anti-digoxigenin antibody attached to a solid support; and detection of the target oligonucleotide using tagged streptavidin or radioactivity measurements.

(iii) separation of the target oligonucleotide which is bound to protein from the target oligonucleotide which is not bound to protein by passing the assay mixture through a nitrocellulose filter under conditions in which the protein:DNA complexes are retained by the nitrocellulose while the non-protein bound DNA passes through the nitrocellulose; and detection of the target oligonucleotide using radioactivity, tagged anti-digoxigenin:digoxigenin interactions, or tagged streptavidin:biotin interactions.

Brief Description of the Figures

Figure 1A illustrates a DNA-binding protein binding to a screening sequence. Figures 1B and 1C illustrate how a DNA-binding protein may be displaced or hindered in binding by a small m lecule by two different mechanisms: because

of steric hinderance (1B) or because of conformational (allosteric) changes induced in the DNA by a small molecule (1C).

Figure 2 illustrates an assay for detecting inhibitory molecules based on their ability to preferentially hinder the binding of a DNA-binding protein to its binding site. Protein (O) is displaced from DNA (/) in the presence of inhibitor (X). Two alternative capture/detection systems are illustrated, the capture and detection of unbound DNA or the capture and detection of DNA:protein complexes.

Figure 3 shows a DNA-binding protein that is able to protect a biotin moiety, covalently attached to the oligonucleotide sequence, from being recognized by the streptavidin when the protein is bound to the DNA.

Figure 4A shows the incorporation of biotin and digoxigenin into a typical oligonucleotide molecule for use in the assay of the present invention. The oligonucleotide contains the binding sequence (i.e., the screening sequence) of the UL9 protein, which is underlined, and test sequences flanking the screening sequence. Figure 4B shows the preparation of double-stranded oligonucleotides end-labeled with either digoxigenin or ^{32}P.

Figure 5 shows a series of sequences that have been tested in the assay of the present invention for the binding of sequence-specific small molecules.

Figure 6 outlines the cloning of a truncated form of the UL9 protein, which retains its sequence-specific DNA-binding ability (UL9-COOH), into an expression vector.

Figure 7 shows the pVL1393 baculovirus vector containing the full length UL9 protein coding sequence.

Figure 8 is a photograph of a SDS-polyacrylamide gel showing (i) the purified UL9-COOH/glutathione-S-transferase fusion protein and (ii) the UL9-COOH polypeptide. In th figure the UL9-COOH polypeptide is indicated by an arrow.

Figure 9 shows the effect on UL9-COOH binding of alterations in the test sequences that flank the UL9 screening sequence. The data are displayed on band shift gels.

Figure 10A shows the effect of the addition of several concentrations of Distamycin A to DNA:protein assay reactions utilizing different test sequences. Figure 10B shows the effect of the addition of Actinomycin D to DNA:protein assay reactions utilizing different test sequences. Figure 10C shows the effect of the addition of Doxorubicin to DNA:protein assay reactions utilizing different test sequences.

Figure 11A illustrates a DNA capture system of the present invention utilizing biotin and streptavidin coated magnetic beads. The presence of the DNA is detected using an alkaline-phosphatase substrate that yields a chemiluminescent product. Figure 11B shows a similar reaction using biotin coated agarose beads that are conjugated to streptavidin, that in turn is conjugated to the captured DNA.

Figure 12 demonstrates a test matrix based on DNA:protein-binding data.

Figure 13 lists the top strands (5′-3′) of all the possible four base pair sequences that could be used as a defined set of ordered test sequences in the assay (for a screening sequence having n bases, where n=4).

Figure 14 lists the top strands (5′-3′) of all the possible four base pair sequences that have the same base composition as the sequence 5′-GATC-3′. This is another example of a defined, ordered set of sequences that could be tested in the assay.

Figure 15 shows an example of an oligonucleotide m lecule containing test sequences flanking a screening sequence. The sequence of this molecule is presented as SEQ ID NO:18, where the "X" of Figure 15 is N in SEQ ID

NO:18.

Detailed Description of the Invention

Definitions:

Adjacent is used to describe the distance relationship between two neighboring DNA sites. Adjacent sites are 20 or less bp apart, or more preferably, 10 or less bp apart, or even more preferably, 5 or less bp apart, or most preferably, immediately abutting one another. "Flanking" is a synonym for adjacent.

Bound DNA, as used in this disclosure, refers to the DNA that is bound by the protein used in the assay (ie., in the examples of this disclosure, the UL9 protein).

Dissociation is the process by which two molecules cease to interact: the process occurs at a fixed average rate under specific physical conditions.

Functional binding is the noncovalent association of a protein or small molecule to the DNA molecule. In the assay of the present invention the functional binding of the protein to the screening sequence (i.e., its cognate DNA binding site) has been evaluated using filter binding or gel band-shift experiments.

Heteromolecules are molecules that are comprised of at least two different types of molecules: for example, the covalent coupling of at least two small organic DNA-binding molecules (eg., distamycin, actinomycin D, or acridine) to each other or the covalent coupling of such a DNA-binding molecule(s) to a DNA-binding polymer (eg., a deoxyoligonucleotide).

On-rate is herein defined as the time required for two molecules to reach steady state association: for example, the DNA:protein complex.

Off-rate is herein d fined as the time required for one-half of the associated complexes, e.g., DNA:protein compl xes, to dissociate.

Sequence-specific binding refers to DNA binding molecules which have a strong DNA sequence binding preference. For example, restriction enzymes and the proteins listed in Table I demonstrate typical sequence-specific DNA-binding.

Sequence-preferential binding refers to DNA binding molecules that generally bind DNA but that show preference for binding to some DNA sequences over others. Sequence-preferential binding is typified by several of the small molecules tested in the present disclosure, e.g., distamycin. Sequence-preferential and sequence-specific binding can be evaluated using a test matrix such as is presented in Figure 12. For a given DNA-binding molecule, there are a spectrum of differential affinities for different DNA sequences ranging from non-sequence-specific (no detectable preference) to sequence preferential to absolute sequence specificity (ie., the recognition of only a single sequence among all possible sequences, as is the case with many restriction endonucleases).

Screening sequence is the DNA sequence that defines the cognate binding site for the DNA binding protein: in the case of UL9 the screening sequence can, for example, be SEQ ID NO:1.

Small molecules are desirable as therapeutics for several reasons related to drug delivery: (i) they are commonly less than 10 K molecular weight; (ii) they are more likely to be permeable to cells; (iii) unlike peptides or oligonucleotides, they are less susceptible to degradation by many cellular mechanisms; and, (iv) they are not as apt to elicit an immune response. Many pharmaceutical companies have extensive libraries of chemical and/or biological mixtures, often fungal, bacterial, or algal extracts, that would be desirable to screen with the assay of the present invention. Small molecules may be either biological or synthetic organic

compounds, or even inorganic compounds (i.e., cisplatin).

Test sequence is a DNA sequence adjacent the screening sequence. The assay of the present invention screens for molecules that, when bound to the test sequence, affect the interaction of the DNA-binding protein with its cognate binding site (i.e., the screening sequence). Test sequences can be placed adjacent either or both ends of the screening sequence. Typically, binding of molecules to the test sequence interferes with the binding of the DNA-binding protein to the screening sequence. However, some molecules binding to these sequences may have the reverse effect, causing an increased binding affinity of the DNA-binding protein to the screening sequence. Some molecules, even while binding in a sequence specific or sequence preferential manner, might have no effect in the assay. These molecules would not be detected in the assay.

Unbound DNA, as used in this disclosure, refers to the DNA that is not bound by the protein used in the assay (ie., in the examples of this disclosure, the UL9 protein).

I. The Assay

One feature of the present invention is that it provides an assay to identify small molecules that will bind in a sequence-specific manner to medically significant DNA target sites. The assay facilitates the development of a new field of pharmaceuticals that operate by interfering with specific DNA functions, such as crucial DNA:protein interactions. A sensitive, well-controlled assay to detect DNA-binding molecules and to determine their sequence-specificity and affinity has been developed. The assay can be used to screen large biological and chemical libraries; for xample, the assay will be used to detect sequence-specific DNA-binding m lecules in fermentation broths or extracts from vari us microorganisms. Furthermore, another applicati n for the assay is to determine the sequence

specificity and relative affinities of *known* DNA-binding drugs (and other DNA-binding molecules) for different DNA sequences. The drugs, which are primarily used in anticancer treatments, may have previously unidentified activities that make them strong candidates for therapeutics or therapeutic precursors in entirely different areas of medicine.

The screening assay is basically a competition assay that is designed to test the ability of a molecule to compete with a DNA-binding protein for binding to a short, synthetic, double-stranded oligodeoxynucleotide that contains the recognition sequence for the DNA-binding protein flanked on either or both sides by a variable test site. The variable test site may contain any DNA sequence that provides a reasonable recognition sequence for a DNA-binding molecule. Molecules that bind to the test site alter the binding characteristics of the protein in a manner that can be readily detected; the extent to which such molecules are able to alter the binding characteristics of the protein is likely to be directly proportional to the affinity of the test molecule for the DNA test site. The relative affinity of a given molecule for different oligonucleotide sequences at the test site (i.e., the test sequences) can be established by examining its effect on the DNA:protein interaction in each of the oligonucleotides. The determination of the high affinity DNA binding sites for DNA-binding molecules will allow us to identify specific target sequences for drug development.

A. General Considerations.

The assay of the present invention has been designed for detecting test molecules or compounds that affect the rat of transfer of a specific DNA molecule from one pr tein m lecul t another identical protein in solution. A mixture of DNA and protein is pr pared in solution.

The concentration of protein is in excess to the concentration of the DNA so that virtually all of the DNA is found in DNA:protein complexes. The DNA is a double-stranded oligonucleotide that contains the recognition sequence for a specific DNA-binding protein (i.e., the screening sequence). The protein used in the assay contains a DNA-binding domain that is specific for binding to the sequence within the oligonucleotide. The physical conditions of the solution (e.g., pH, salt concentration, temperature) are adjusted such that the half-life of th complex is amenable to performing the assay (optimally a half-life of 5-30 minutes), preferably in a range that is close to normal physiological conditions.

As one DNA:protein complex dissociates, the released DNA rapidly reforms a complex with another protein in solution. Since the protein is in excess to the DNA, dissociations of one complex always result in the rapid reassociation of the DNA into another DNA:protein complex. At equilibrium, very few DNA molecules will be unbound. The minimum background of the assay is the amount of unbound DNA observed during any given measurable time period. The brevity of the observation period and the sensitivity of the detection system define the lower limits of background DNA.

Figure 1 illustrates how such a protein can be displaced from its cognate binding site or how a protein can be prevented from binding its cognate binding site, or how the kinetics of the DNA:protein interaction can b altered. One mechanism is steric hinderance of protein binding by a small molecule. Alternatively, a molecule may interfere with a DNA:protein binding interaction by inducing a conformational change in the DNA. In either event, if a test molecule that binds the oligonucleotide hinders binding of the protein, the rate of transfer of DNA from one pr tein to another will be decreased. This will

result in a net increase in the amount of unbound DNA. In other words, an increase in the amount of unbound DNA or a decrease in the amount of bound DNA indicates the presence of an inhibitor.

Alternatively, molecules may be isolated that, when bound to the DNA, cause an increased affinity of the DNA-binding protein for its cognate binding site. In this case the amount of unbound DNA (observed during a given measurable time period after the addition of the molecule) will decrease in the reaction mixture as detected by the capture/detection system described in Section II.

B. Other Methods

There are several approaches that could be taken to look for small molecules that specifically inhibit the interaction of a given DNA-binding protein with its binding sequence (cognate site). One approach would be to test biological or chemical compounds for their ability to preferentially block the binding of one specific DNA:protein interaction but not the others. Such an assay would depend on the development of at least two, preferably three, DNA:protein interaction systems in order to establish controls for distinguishing between general DNA-binding molecules (polycations like heparin or intercalating agents like ethidium) and DNA-binding molecules having sequence binding preferences that would affect protein/cognate binding site interactions in one system but not the other(s).

One illustration of how this system could be used is as follows. Each cognate site could be placed 5′ to a reporter gene (such as genes encoding β-galactoside or luciferase) such that binding of the protein to the cognate site would enhance transcripti n f the reporter gene. The presence of a sequence-specific DNA-binding drug that blocked the DNA:pr tein interaction would decrease th

enhancement of the reporter gene expression. Several DNA enhancers could be coupled to reporter genes, then each construct compared to one another in the presence or absence of small DNA-binding test molecules. In the case where multiple protein/cognate binding sites are used for screening, a competitive inhibitor that blocks one interaction but not the others could be identified by the lack of transcription of a reporter gene in a transfected cell line or in an *in vitro* assay. Only one such DNA-binding sequence, specific for the protein of interest, could be screened with each assay system. This approach has a number of limitations including limited testing capability and the need to construct the appropriate reporter system for each different protein/cognate site of interest.

C. Choosing and Testing an Appropriate DNA-Binding Protein.

Experiments performed in support of the present invention have defined a second approach for identifying molecules having sequence-preferential DNA-binding. In this approach small molecules binding to sequences adjacent the cognate binding sequence can inhibit the protein/cognate DNA interaction. This assay has been designed to use a single DNA:protein interaction to screen for sequence-specific or sequence-preferential DNA-binding molecules that recognize virtually any sequence.

While DNA-binding recognition sites are usually quite small (4-17 bp), the sequence that is protected by the binding protein is larger (usually 5 bp or more on either side of the recognition sequence -- as detected by DNAase I protection (Galas et al.) or methylation interference (Siebenlist et al.). Experiments performed in support of th present invention demonstrated that a single protein and its cognate DNA-binding sequence can be used to assay virtually any DNA sequence by placing a sequenc of

interest adjacent to the cognate site: a small molecule bound to the adjacent site can be detected by alterations in the binding characteristics of the protein to its cognate site. Such alterations might occur by either steric hindrance, which would cause the dissociation of the protein, or induced conformational changes in the recognition sequence for the protein, which may cause either enhanced binding or more likely, decreased binding of the protein to its cognate site.

1) Criteria for choosing an appropriate DNA-binding protein.

There are several considerations involved in choosing DNA:protein complexes that can be employed in the assay of the present invention including:

a) The off-rate (see "Definitions") should be fast enough to accomplish the assay in a reasonable amount of time. The interactions of some proteins with cognate sites in DNA can be measured in days not minutes: such tightly bound complexes would inconveniently lengthen the period of time it takes to perform the assay.

b) The off-rate should be slow enough to allow the measurement of unbound DNA in a reasonable amount of time. For example, the level of free DNA is dictated by the ratio between the time needed to measure free DNA and the amount of free DNA that occurs naturally due to the off-rate during the measurement time period.

In view of the above two considerations, practical useful DNA:protein off-rates fall in the range of approximately two minutes to several days, although shorter off-rates may be accomodated by faster equipment and longer off-rates may be accomodated by destabilizing the binding conditions for the assay.

c) A further consideration is that the kinetic interactions of the DNA:protein complex is relatively insensitive to th nucleotide sequences flanking th

recognition sequence. The affinity of many DNA-binding proteins is affected by differences in the sequences adjacent to the recognition sequence. The most obvious example of this phenomenon is the preferential binding and cleavage of restriction enzymes given a choice of several identical recognition sequences with different flanking sequences (Polinsky et al.). If the off-rates are affected by flanking sequences the analysis of comparative binding data between different flanking oligonucleotide sequences becomes difficult but is not impossible.

2) Testing DNA:protein interactions for use in the assay.

Experiments performed in support of the present invention have identified a DNA:protein interaction that is particularly useful for the above described assay: the Herpes Simplex Virus (HSV) UL9 protein that binds the HSV origin of replication (*oriS*). The UL9 protein has fairly stringent sequence specificity. There appear to be three binding sites for UL9 in *oriS*, SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:17 (Elias, P. et al., Stow et al.). One sequence (SEQ ID NO:1) binds with at least 10-fold higher affinity than the second sequence (SEQ ID NO:2): the embodiments described below use the higher affinity binding site (SEQ ID NO:1).

DNA:protein association reactions are performed in solution. The DNA:protein complexes can be separated from free DNA by any of several methods. One particularly useful method for the initial study of DNA:protein interactions has been visualization of binding results using band shift gels (Example 3A). In this method DNA:protein binding reactions are applied to polyacrylamide/TBE gels and the labelled complexes and free labeled DNA ar separated electroph retically. These gels are fixed, dried, and exp sed to X-ray film. The resulting autoradiograms are examined for the amount of free probe

that is migrating separately from the DNA:protein complex. These assays include (i) a lane containing only free labeled probe, and (ii) a lane where the sample is labeled probe in the presence of a large excess of binding protein. The band shift assays allow visualization of the ratios between DNA:protein complexes and free probe. However, they are less accurate than filter binding assays for rate-determining experiments due to the lag time between loading the gel and electrophoretic separation of the components.

The filter binding method is particularly useful in determining the off-rates for protein:oligonucleotide complexes (Example 3B). In the filter binding assay, DNA:protein complexes are retained on a filter while free DNA passes through the filter. This assay method is more accurate for off-rate determinations because the separation of DNA:protein complexes from free probe is very rapid. The disadvantage of filter binding is that the nature of the DNA:protein complex cannot be directly visualized. So if, for example, the competing molecule was also a protein competing for the binding of a site on the DNA molecule, filter binding assays cannot differentiate between the binding of the two proteins nor yield information about whether one or both proteins are binding.

There are many known DNA:protein interactions that may be useful in the practice of the present invention, including (i) the DNA protein interactions listed in Table I, (ii) bacterial, yeast, and phage systems such as lambda O_L-O_R/cro, and (iii) modified restriction enzyme systems (e.g., protein binding in the absence of divalent cations). Any protein that binds to a specific recognition sequence may be useful in the present invention. One constraining factor is the effect of the immediately adjacent sequences (the test sequences) on the affinity of the protein for its recognition sequence. DNA:protein interactions in which

there is little or no effect of the test sequences on the affinity of the protein for its cognate site are preferable for use in the described assay; however, DNA:protein interactions that exhibit (test sequence-dependent) differential binding may still be useful if algorithms are applied to the analysis of data that compensate for the differential affinity. In general, the effect of flanking sequence composition on the binding of the protein is likely to be correlated to the length of the recognition sequence for the DNA-binding protein. In short, the kinetics of binding for proteins with shorter recognition sequences are more likely to suffer from flanking sequenc effects, while the kinetics of binding for proteins with longer recognition sequences are more likely to not be affected by flanking sequence composition. The present disclosure provides methods and guidance for testing the usefulness of such DNA:protein interactions, i.e., other than the UL9 oriS binding site interaction, in the screening assay.

D. Preparation of Full Length UL9 and UL9-COOH Polypeptides.

UL9 protein has been prepared by a number of recombinant techniques (Example 2). The full length UL9 protein has been prepared from baculovirus infected insect cultures (Example 3A, B, and C). Further, a portion of the UL9 protein that contains the DNA-binding domain (UL9-COOH) has been cloned into a bacterial expression vector and produced by bacterial cells (Example 3D and E). The DNA-binding domain of UL9 is contained within the C-terminal 317 amino acids of the protein (Weir et al.). The UL9-COOH polypeptide was inserted into the expression vect r in-frame with the glutathione-S-transferase (gst) protein. The gst/UL9 fusion protein was purified using affinity chromatography (Example 3E). The vector also contained a thrombin cleavage sit at the junction of the two

polypeptides. Therefore, once the fusion protein was isolated (Figure 8, lane 2) it was treated with thrombin, cleaving the UL9-COOH/gst fusion protein from the gst polypeptide (Figure 8, lane 3). The UL9-COOH-gst fusion polypeptide was obtained at a protein purity of greater than 95% as determined using Coomaisie staining.

Other hybrid proteins can be utilized to prepare DNA-binding proteins of interest. For example, fusing a DNA-binding protein coding sequence in-frame with a sequence encoding the thrombin site and also in-frame with the β-galactoside coding sequence. Such hybrid proteins can be isolated by affinity or immunoaffinity columns (Maniatis et al.; Pierce, Rockford IL). Further, DNA-binding proteins can be isolated by affinity chromatography based on their ability to interact with their cognate DNA binding site. For example, the UL9 DNA-binding site (SEQ ID NO:1) can be covalently linked to a solid support (e.g., CnBr-activated Sepharose 4B beads, Pharmacia, Piscataway NJ), extracts passed over the support, the support washed, and the DNA-binding then isolated from the support with a salt gradient (Kadonaga). Alternatively, other expression systems in bacteria, yeast, insect cells or mammalian cells can be used to express adequate levels of a DNA-binding protein for use in this assay.

The results presented below in regard to the DNA-binding ability of the truncated UL9 protein suggest that full length DNA-binding proteins are not required for the DNA:protein assay of the present invention: only a portion of the protein containing the cognate site recognition function may be required. The portion of a DNA-binding protein required for DNA-binding can be evaluated using a functional binding assay (Example 4A). The rate of dissociation can be evaluated (Example 4B) and compared to that of the full length DNA-binding protein. However, any DNA-binding peptide, truncated or full length, may be used

in the assay if it meets the criteria outlined in part I.C.1, "Criteria for choosing an appropriate DNA-binding protein". This remains true whether or not the truncated form of the DNA-binding protein has the same affinity as the full length DNA-binding protein.

E. Functional Binding and Rate of Dissociation.

The full length UL9 and purified UL9-COOH proteins were tested for functional activity in "band shift" assays (see Example 4A). The buffer conditions were optimized for DNA:protein-binding (Example 4C) using the UL9-COOH polypeptide. These DNA-binding conditions also worked well for the full-length UL9 protein. Radiolabelled oligonucleotides (SEQ ID NO:14) that contained the 11 bp UL9 DNA-binding recognition sequence (SEQ ID NO:1) were mixed with each UL9 protein in appropriate binding buffer. The reactions were incubated at room temperature for 10 minutes (binding occurs in less than 2 minutes) and the products were separated electrophoretically on non-denaturing polyacrylamide gels (Example 4A). The degree of DNA:protein-binding could be determined from the ratio of labeled probe present in DNA:protein complexes versus that present as free probe. This ratio was typically determined by optical scanning of autoradiograms and comparison of band intensities. Other standard methods may be used as well for this determination, such as scintillation counting of excised bands. The UL9-COOH polypeptide and the full length UL9 polypeptide, in their respective buffer conditions, bound the target oligonucleotide equally well.

The rate of dissociation was determined using competition assays. An excess of unlabelled oligonucleotide that contained the UL9 binding site was added to each reaction. This unlabelled oligonucleotid acts as a specific inhibitor, capturing the UL9 protein as it dissociates from the labelled ligonucleotide (Example

4B). The dissociation rate, as determined by a band-shift assay, for both full length UL9 and UL9-COOH was approximately 4 hours at 4°C or approximately 10 minutes at room temperature. Neither non-specific oligonucleotides (a 10,000-fold excess) nor sheared herring sperm DNA (a 100,000-fold excess) competed for binding with the oligonucleotide containing the UL9 binding site.

F. *oriS* Flanking Sequence Variation.

As mentioned above, one feature of a DNA:protein-binding system for use in the assay of the present invention is that the DNA:protein interaction is not affected by the nucleotide sequence of the regions adjacent the DNA-binding site. The sensitivity of any DNA:protein-binding reaction to the composition of the flanking sequences can be evaluated by the functional binding assay and dissociation assay described above.

To test the effect of flanking sequence variation on UL9 binding to the *oriS* SEQ ID NO:1 sequences oligonucleotides were constructed with 20-30 different sequences (i.e., the test sequences) flanking the 5′ and 3′ sides of the UL9 binding site. Further, oligonucleotides were constructed with point mutations at several positions within the UL9 binding site. Most point mutations within the binding site destroyed recognition. Several changes did not destroy recognition and these include variations at sites that differ between the three UL9 binding sites (SEQ ID NO:1, SEQ ID NO:2 and SEQ ID NO:17): the second UL9 binding site (SEQ ID NO:2) shows a ten-fold decrease in UL9:DNA binding affinity (Elias et al.) relative to the first (SEQ ID NO:1). On the other hand, sequence variation at the test site (also called the test sequence), adjacent to the screening site (Figure 5, Exampl 5), had virtually no effect n binding or th rate of dissociation.

Th results demonstrating that the nucleotide sequence

in the test site, which flanks the screening site, has no effect on the kinetics of UL9 binding in any of the oligonucleotides tested is a striking result. This allows the direct comparison of the effect of a DNA-binding molecule on test oligonucleotides that contain different test sequences. Since the only difference between test oligonucleotides is the difference in nucleotide sequence at the test site(s), and since the nucleotide sequence at the test site has no effect on UL9 binding, any differential effect observed between the two test oligonucleotides in response to a DNA-binding molecule must be due solely to the differential interaction of the DNA-binding molecule with the test sequence(s). In this manner, the insensitivity of UL9 to the test sequences flanking the UL9 binding site greatly facilitates the interpretation of results. Each test oligonucleotide acts as a control sample for all other test oligonucleotides. This is particularly true when ordered sets of test sequences are tested (eg., testing all 256 four base pair sequences (Figure 13) for binding to a single drug).

Taken together the above experiments support that the UL9-COOH polypeptide binds the SEQ ID NO:1 sequence with (i) appropriate strength, (ii) an acceptable dissociation time, and (iii) indifference to the nucleotide sequences flanking the assay (binding) site. These features suggested that the UL9/*oriS* system could provide a versatile assay for detection of small molecule/DNA-binding involving any number of specific nucleotide sequences.

The above-described experiment can be used to screen other DNA:protein interactions to determine their usefulness in the present assay.

G. Small Molecules as Sequence-Specific Competitive Inhibitors.

To test the utility of the present assay system

several small molecules that have sequence preferences (e.g., a preference for AT-rich versus GC-rich sequences) have been tested.

Distamycin A binds relatively weakly to DNA (K_A = 2 x 10^5 M^{-1}) with a preference for non-alternating AT-rich sequences (Jain et al.; Sobell; Sobell et al.). Actinomycin D binds DNA more strongly (K_A = 7.6 x 10^{-7} M^{-1}) than Distamycin A and has a relatively strong preference for the dinucleotide sequence dGdC (Luck et al.; Zimmer; Wartel). Each of these molecules poses a stringent test for the assay. Distamycin A tests the sensitivity of the assay because of its relatively weak binding. Actinomycin D challenges the ability to utilize flanking sequences since the UL9 recognition sequence contains a dGdC dinucleotide: therefore, it might be anticipated that all of the oligonucleotides, regardless of the test sequence flanking the assay site, might be equally affected by actinomycin D.

In addition, Doxorubicin, a known anti-cancer agent that binds DNA in a sequence-preferential manner (Chen, K-X, et al.), has been tested for preferential DNA sequence binding using the assay of the present invention.

Actinomycin D, Distamycin A, and Doxorubicin have been tested for their ability to preferentially inhibit the binding of UL9 to oligonucleotides containing different sequences flanking the UL9 binding site (Example 6, Figure 5). Binding assays were performed as described in Example 5. These studies were completed under conditions in which UL9 is in excess of the DNA (i.e., most of the DNA is in complex).

Distamycin A was tested with 5 different test sequences flanking the UL9 screening sequence: SEQ ID NO:5 to SEQ ID NO:9. The results shown in Figure 10A demonstrat that distamycin A preferentially disrupts

binding to the test sequences UL9 polyT, UL9 polyA and, to a lesser extent, UL9 ATAT. Figure 10A also shows the concentration dependence of the inhibitory effect of distamycin A: at 1 μM distamycin A most of the DNA:protein complexes are intact (top band) with free probe appearing in the UL9 polyT and UL9 polyA lanes, and some free probe appearing in the UL9 ATAT lane; at 4 μM free probe can be seen in the UL9 polyT and UL9 polyA lanes; at 16 μM free probe can be seen in the UL9 polyT and UL9 polyA lanes; and at 40 μM the DNA:protein in the polyT, UL9 polyA and UL9 ATAT lanes are near completely disrupted while some DNA:protein complexes in the other lanes persist. These results are consistent with Distamycin A's known binding preference for non-alternating AT-rich sequences.

Actinomycin D was tested with 8 different test sequences flanking the UL9 screening sequence: SEQ ID NO:5 to SEQ ID NO:9, and SEQ ID NO:11 to SEQ ID NO:13. The results shown in Figure 10B demonstrate that actinomycin D preferentially disrupts the binding of UL9-COOH to the oligonucleotides UL9 CCCG (SEQ ID NO:5) and UL9 GGGC (SEQ ID NO:6). These oligonucleotides contain, respectively, three or five dGdC dinucleotides in addition to the dGdC dinucleotide within the UL9 recognition sequence. This result is consistent with Actinomycin D's known binding preference for the dinucleotide sequence dGdC. Apparently the presence of a potential target site within the screening sequence (*oriS*, SEQ ID NO:1), as mentioned above, does not interfere with the function of the assay.

Doxorubicin was tested with 8 different test sequences flanking the UL9 screening sequence: SEQ ID NO:5 to SEQ ID NO:9, and SEQ ID NO:11 to SEQ ID NO:13. The results shown in Figure 10C demonstrate that Doxorubicin preferentially disrupts binding to oriEco3, the test sequence of which differs from oriEco2 by only one base (compare SEQ ID NO:12 and SEQ ID NO:13). Figure 10C also shows the concentration

dependence of the inhibitory effect of Doxorubicin: at 15 μM Doxorubicin, the UL9 binding to the screening sequence is strongly affected when oriEco3 is the test sequence, and more mildly affected when polyT, UL9 GGGC, or oriEco2 was the test sequence; and at 35 μM Doxorubicin most DNA:protein complexes are nearly completely disrupted, with UL9 polyT and UL9ATAT showing some DNA still complexed with protein. Also, effects similar to those observed at 15 μM were also observed using Doxorubicin at 150 nM, but at a later time point.

Further incubation with any of the drugs resulted in additional disruption of binding. Given that the one hour incubation time of the above assays is equivalent to several half-lives of the DNA:protein complex, the additional disruption of binding suggests that the on-rate for the drugs is comparatively slow.

The ability of the assay to distinguish sequence binding preference using weak DNA-binding molecules with poor sequence-specificity (such as distamycin A) is a stringent test. Accordingly, the present assay seems well-suited for the identification of molecules having better sequence specificity and/or higher sequence binding affinity. Further, the results demonstrate sequence preferential binding with the known anti-cancer drug Doxorubicin. This result indicates the assay may be useful for screening mixtures for molecules displaying similar characteristics that could be subsequently tested for anti-cancer activities as well as sequence-specific binding.

Other compounds that may be suitable for testing the present DNA:protein system or for defining alternate DNA:protein systems include the following: echinomycin, which preferentially binds to the sequence (A/T)CGT (Quigley et al.); small inorganic molecules, such as cobalt hexamine, that are known to induce Z-DNA formation in regions that contain repetitive GC sequences (Gessner et

al.); and other DNA-binding proteins, such as EcoR1, a restriction endonuclease.

H. Theoretical considerations on the concentration of assay components.

There are two components in the assay, the test sequence (oligonucleotide) and the DNA-binding domain of UL9, which is described below. A number of theoretical considerations have been employed in establishing the assay system of the present invention. In one embodiment of the invention, the assay is used as a mass-screening assay. In this capacity, small volumes and concentrations were desirable. A typical assay uses about 0.1 ng DNA in a 15-20 µl reaction volume (approximately 0.3 nM). The protein concentration is in excess and can be varied to increase or decrease the sensitivity of the assay. In the simplest scenario, where the small molecule is acting as a competitive inhibitor via steric hindrance, the system kinetics can be described by the following equations:

$$D + P \rightleftharpoons D:P, \text{ where } k_{fp}/k_{bp} = K_{eq,p} = [D:P]/[D][P]$$

and

$$D + X \rightleftharpoons D:X, \text{ where } k_{fx}/k_{bx} = K_{eq,x} = [D:X]/[D][X]$$

D = DNA, P = protein, X = DNA-binding molecule, k_{fp} and k_{fx} are the rates of the forward reaction for the DNA:protein interaction and DNA:drug interaction, respectively, and k_{bp} and k_{bx} are the rates of the backwards reactions for the respective interactions. Brackets, [], indicate molar concentration of the components.

In the assay, both the protein, P, and the DNA-binding molecule or drug, X, are competing for the DNA. If steric hindrance is the mechanism of inhibition, the assumption can be made that the two molecules are competing for the same site. When the concentration of DNA equals the concentration of the DNA:drug or DNA:protein complex, the equilibrium binding constant, K_{eq}, is equal to the reciprocal of the protein concentration (1/[P]). For UL9, the calculated $K_{eq,UL9} = 2.2 \times 10^9\ M^{-1}$. When all three components are mixed together, the relationship between the drug and the protein can be described as:

$$K_{eq,P} = z(K_{eq,X})$$

where "z" defines the difference in affinity for the DNA between P and X. For example, if z =4, then the affinity of the drug is 4-fold lower than the affinity of the protein for the DNA molecule. The concentration of X, therefore, must be 4-fold greater than the concentration of P, to compete equally for the DNA molecule. Thus, the equilibrium affinity constant of UL9 will define the minimum level of detection with respect to the concentration and/or affinity of the drug. Low affinity DNA-binding molecules will be detected only at high concentrations; likewise, high affinity molecules can be detected at relatively low concentrations.

With certain test sequences, complete inhibition of UL9 binding at markedly lower concentrations than indicated by these analyses have been observed, probably indicating that certain sites among those chosen for feasibility studies have affinities higher than previously published. Note that relatively high concentrations of known drugs can be utilized for t sting sequence specificity. In addition, the binding constant of UL9 can be readily lowered by

altering the pH or salt concentration in the assay if it is desirable to screen for molecules that are found at low concentration (eg., in a fermentation broth or extract).

Analyses such as presented above, become more complex if the inhibition is allosteric (non-competitive inhibition) rather than competition by steric hindrance. Nonetheless, the probability that the relative effect of an inhibitor on different test sequences is due to its relative and differential affinity to the different test sequences is fairly high. This is particularly true in the assays in which all sequences within an ordered set (eg., possible sequences of a given length or all possible variations of a certain base composition and defined length) are tested. In brief, if the effect of inhibition in the assay is particularly strong for a single sequence, then it is likely that the inhibitor binds that particular sequence with higher affinity than any of the other sequences. Furthermore, while it may be difficult to determine the absolute affinity of the inhibitor, the relative affinities have a high probability of being reasonably accurate. This information will be most useful in facilitating, for instance, the refinement of molecular modeling systems.

I. The use of the assay under conditions of high protein concentration.

When the screening protein is added to the assay system at very high concentrations, the protein binds to non-specific sites on the oligonucleotide in addition to the screening sequence. This effect has been demonstrated using band shift gels: in particular, when serial dilutions are made of the UL-9 protein and the dilutions are mixed with a fixed concentration of oligonucleotide, no binding (as seen by a band shift) is observed at very low dilutions (e.g., 1:100,000), a single band shift is observed at mod rate dilutions (e.g., 1:100) and a smear, migrating

higher than the single band observed at moderated dilutions, is observed at high concentrations of protein (e.g., 1:10). In the band shift assay, a smear is indicative of a mixed population of complexes, all of which presumably have the screening protein binding to the screening sequence with high affinity (e.g., for UL9, K_2 = 1.1×10^9 M^{-1}) but in addition have a larger number of proteins bound with markedly lower affinity.

Some of the low affinity binding proteins are bound to the test sequence. In experiments performed in support of the present invention, using mixtures of UL9 and glutathione-S-transferase, the low affinity binding proteins are likely UL9 or, less likely, glutathione-S-transferase, since these are the only proteins in the assay mixture. These low affinity binding proteins are significantly more sensitive to interference by a molecule binding to the test sequence for two reasons. First, the interference is likely to be by direct steric hinderance and does not rely on induced conformational changes in the DNA; secondly, the protein binding to the test site is a low affinity binding protein because the test site is not a cognate-binding sequence. In the case of UL9, the difference in affinity between the low affinity binding and the high affinity binding appears to be at least two orders of magnitude.

Experiments performed in support of the present invention demonstrate that the filter binding assays capture more DNA:protein complexes when more protein is bound to the DNA. The relative results are accurate, but under moderate protein concentrations, not all of the bound DNA (as demonstrated by band shift assays) will bind to the filter unless there is more than one DNA:protein complex per oligonucleotide (e.g., in the case of UL9, more than one UL9:DNA complex). This makes the assay exquisit ly

sensitive under conditions of high protein concentration. For instance, when actinomycin binds DNA at a test site under conditions where there is one DNA:UL9 complex per oligonucleotide, a differential-binding effect on GC-rich oligonucleotides has been observed (see Example 6). Under conditions of high protein concentration, where more than one DNA:UL9 complex is found per oligonucleotide, the differential effect of actinomycin D is even more marked. These results suggest that the effect of actinomycin D on a test site that is weakly bound by protein may be more readily detected than the effect of actinomycin D on the adjacent screening sequence. Therefore, employing high protein concentrations may increase the sensitivity of the assay.

II. Capture/Detection Systems.

As an alternative to the above described band shift gels and filter binding assays, the measurement of inhibitors can be monitored by measuring either the level of unbound DNA in the presence of test molecules or mixtures or the level of DNA:protein complex remaining in the presence of test molecules or mixtures. Measurements may be made either at equilibrium or in a kinetic assay, prior to the time at which equilibrium is reached. The type of measurement is likely to be dictated by practical factors, such as the length of time to equilibrium, which will be determined by both the kinetics of the DNA:protein interaction as well as the kinetics of the DNA:drug interaction. The results (ie., the detection of DNA-binding molecules and/or the determination of their sequence preferences) should not vary with the type of measurement taken (kinetic or equilibrium).

Figure 2 illustrates an assay for det cting inhibitory molecules bas d on th ir ability to preferentially hinder the binding of a DNA-binding prot in. In the presence of

an inhibitory molecule (X) the equilibrium between the DNA-binding protein and its binding site (screening sequence) is disrupted. The DNA-binding protein (O) is displaced from DNA (/) in the presence of inhibitor (X), the DNA free of protein or, alternatively, the DNA:protein complexes, can then be captured and detected.

For maximum sensitivity, unbound DNA and DNA:protein complexes should be sequestered from each other in an efficient and rapid manner. The method of DNA capture should allow for the rapid removal of the unbound DNA from he protein-rich mixture containing the DNA:protein complexes.

Even if the test molecules are specific in their interaction with DNA they may have relatively low affinity and they may also be weak binders of non-specific DNA or have non-specific interactions with DNA at low concentrations. In either case, their binding to DNA may only be transient, much like the transient binding of the protein in solution. Accordingly, one feature of the assay is to take a molecular snapshot of the equilibrium state of a solution comprised of the target/assay DNA, the protein, and the inhibitory test molecule. In the presence of an inhibitor, the amount of DNA that is not bound to protein will be greater than in the absence of an inhibitor. Likewise, in the presence of an inhibitor, the amount of DNA that is bound to protein will be lesser than in the absence of an inhibitor. Any method used to separate the DNA:protein complexes from unbound DNA, should be rapid, because when the capture system is applied to the solution (if the capture system is irreversible), the ratio of unbound DNA to DNA:protein complex will change at a predetermined rate, based purely on the off-rate of the DNA:protein complex. This step, therefore, determines th limits of background. Unlike the protein and inhibitor, the capture system should bind rapidly and tightly to the

DNA or DNA:protein complex. The longer the capture system is left in contact with the entire mixture of unbound DNA and DNA:protein complexes in solution, the higher the background, regardless of the presence or absence of inhibitor.

Two exemplary capture systems are described below for use in the present assay. One capture system has been devised to capture unbound DNA (part II.A). The other has been devised to capture DNA:protein complexes (part II.B). Both systems are amenable to high throughput screening assays. The same detection methods can be applied to molecules captured using either capture system (part II.C)

A. Capture of unbound DNA.

One capture system that has been developed in the course of experiments performed in support of the present invention utilizes a streptavidin/biotin interaction for the rapid capture of unbound DNA from the protein-rich mixture, which includes unbound DNA, DNA:protein complexes, excess protein and the test molecules or test mixtures. Streptavidin binds with extremely high affinity to biotin ($K_d = 10^{-15}M$) (Chaiet et al.; Green), thus two advantages of the streptavidin/biotin system are that binding between the two molecules can be rapid and the interaction is the strongest known non-covalent interaction.

In this detection system a biotin molecule is covalently attached in the oligonucleotide screening sequence (i.e., the DNA-binding protein's binding site). This attachment is accomplished in such a manner that the binding of the DNA-binding protein to the DNA is not destroyed. Further, when the protein is bound to the biotinylated sequence, the protein prevents the binding of streptavidin to the biotin. In other words, the DNA-binding protein is able to protect th biotin from being recognized by the streptavidin. This DNA:protein

interaction is illustrated in Figure 3.

The capture system is described herein for use with the UL9/*oriS* system described above. The following general testing principles can, however, be applied to analysis of other DNA:protein interactions. The usefulness of this system depends on the biophysical characteristics of the particular DNA:protein interaction.

1) Modification of the protein recognition sequence with biotin.

The recognition sequence for the binding of the UL9 (Koff et al.) protein is underlined in Figure 4. Oligonucleotides were synthesized that contain the UL9 binding site and site-specifically biotinylated a number of locations throughout the binding sequence (SEQ ID NO:14; Example 1, Figure 4). These biotinylated oligonucleotides were then used in band shift assays to determine the ability of the UL9 protein to bind to the oligonucleotide. These experiments using the biotinylated probe and a non-biotinylated probe as a control demonstrate that the presence of a biotin at the #8-T (biotinylated deoxyuridine) position of the bottom strand meets the requirements listed above: the presence of a biotin moiety at the #8 position of the bottom strand does not markedly affect the specificity of UL9 for the recognition site; further, in the presence of bound UL9, streptavidin does not recognize the presence of the biotin moiety in the oligonucleotide. Biotinylation at other A or T positions did not have the two necessary characteristics (i.e., UL9 binding and protection from streptavidin): biotinylation at the adenosine in position #8, of the top strand, prevented the binding of UL9; biotinylation of either adenosines or thymidines (top or bottom strand) at positions #3, #4, #10, or #11 all allowed binding f UL9, but in each case, streptavidin also was able to recognise the presence of the biotin moiety and thereby bind the

oligonucleotide in the presence of UL9.

The above result (the ability of UL9 to bind to an oligonucleotide containing a biotin within the recognition sequence and to protect the biotin from streptavidin) was unexpected in that methylation interference data (Koff et al.) suggest that methylation of the deoxyguanosine residues at positions #7 and #9 of the recognition sequence (on either side of the biotinylated deoxyuridine) blocks UL9 binding. In these methylation interference experiments, guanosines are methylated by dimethyl sulfate at the N^7 position, which corresponds structurally to the 5-position of the pyrimidine ring at which the deoxyuridine is biotinylated. These moieties all protrude into the major groove of the DNA. The methylation interference data suggest that the #7 and #9 position deoxyguanosines are contact points for UL9, it was therefore unexpected that the presence of a biotin moiety between them would not interfere with binding.

The binding of the full length protein was relatively unaffected by the presence of a biotin at position #8 within the UL9 binding site. The rate of dissociation was similar for full length UL9 with both biotinylated and un-biotinylated oligonucleotides. However, the rate of dissociation of the truncated UL9-COOH polypeptide was faster with the biotinylated oligonucleotides than with non-biotinylated oligonucleotides, which is a rate comparable to that of the full length protein with either DNA.

The binding conditions were optimized for UL9-COOH so that the off-rate of the truncated UL9 from the biotinylated oligonucleotide was 5-10 minutes (optimized conditions are given in Example 4), a rate compatible with a mass screening assay. The use of multi-well plates to conduct th DNA:protein assay of the present invention is

one approach to mass screening.

2) Capture of site-specific biotinylated oligonucleotides.

The streptavidin:biotin interaction can be employed in several different ways to remove unbound DNA from the solution containing the DNA, protein, and test molecule or mixture. Magnetic polystyrene or agarose beads, to which streptavidin is covalently attached or attached through a covalently attached biotin, can be exposed to the solution for a brief period, then removed by use, respectively, of magnets or a filter mesh. Magnetic streptavidinated beads are currently the method of choice. Streptavidin has been used in many of these experiments, but avidin is equally useful.

An example of a second method for the removal of unbound DNA is to attach streptavidin to a filter by first linking biotin to the filter, binding streptavidin, then blocking nonspecific protein binding sites on the filter with a nonspecific protein such as albumin. The mixture is then passed through the filter, unbound DNA is captured and the bound DNA passes through the filter.

One convenient method to sequester captured DNA is the use of streptavidin-conjugated superparamagnetic polystyrene beads as described in Example 7. These beads are added to the assay mixture to capture the unbound DNA. After capture of DNA, the beads can be retrieved by placing the reaction tubes in a magnetic rack, which sequesters the beads on the reaction chamber wall while the assay mixture is removed and the beads are washed. The captured DNA is then detected using one of several DNA detection systems, as described below.

Alternatively, avidin-coated agarose beads can be used. Biotinylated agarose beads (immobilized D-biotin, Pierce) are bound to avidin. Avidin, like str ptavidin, has four binding sit s for biotin. One of these binding

sites is used to bind the avidin to the biotin that is coupled to the agarose beads via a 16 atom spacer arm: the other biotin binding sites remain available. The beads are mixed with binding mixtures to capture biotinylated DNA (Example 7). Alternative methods (Harlow et al.) to the bead capture methods just described include the following streptavidinated or avidinated supports: low-protein-binding filters or 96-well plates.

B) Capture of DNA:protein complexes.

The amount of DNA:protein complex remaining in the assay mixture in the presence of an inhibitory molecule can also be determined as a measure of the relative effect of the inhibitory molecule. A net decrease in the amount of DNA:protein complex in response to a test molecule is an indication of the presence of an inhibitor. DNA molecules that are bound to protein can be captured on nitrocellulose filters. Under low salt conditions, DNA that is not bound to protein freely passes through the filter. Thus, by passing the assay mixture rapidly through a nitrocellulose filter, the DNA:protein complexes and unbound DNA molecules can be rapidly separated. This has been accomplished on nitrocellulose discs using a vacuum filter apparatus or on slot blot or dot blot apparatuses (all of which are available from Schleicher and Schuell, Keene, NH). The assay mixture is applied to and rapidly passes through the wetted nitrocellulose under vacuum conditions. Any apparatus employing nitrocellulose filters or other filters capable of retaining protein while allowing free DNA to pass through the filter are suitable for this system.

C) Detection systems.

For either of the above capture methods, the amount of DNA that has been captured is quantitated. The method of quantitation d pends on how the DNA has been prepared. If the DNA is radioactively labelled, beads can be counted in a scintillation count r, or autoradi graphs can be taken of

dried gels or nitrocellulose filters. The amount of DNA has been quantitated in the latter case by a densitometer (Molecular Dynamics, Sunnyvale, CA); alternatively, filters or gels containing radiolabeled samples can be quantitated using a phosphoimager (Molecular Dynamics). The captured DNA may be also be detected using a chemiluminescent or colorimetric detection system.

Radiolabelling and chemiluminescence (i) are very sensitive, allowing the detection of sub-femtomole quantities of oligonucleotide, and (ii) use well-established techniques. In the case of chemiluminescent detection, protocols have been devised to accommodate the requirements of a mass-screening assay. Non-isotopic DNA detection techniques have principally incorporated alkaline phosphatase as the detectable label given the ability of the enzyme to give a high turnover of substrate to product and the availability of substrates that yield chemiluminescent or colored products.

1) Radioactive labeling.

Many of the experiments described above for UL9 DNA:protein-binding studies have made use of radio-labelled oligonucleotides. The techniques involved in radiolabelling of oligonucleotides have been discussed above. A specific activity of 10^8-10^9 dpm per μg DNA is routinely achieved using standard methods (eg., end-labeling the oligonucleotide with adenosine γ-[^{32}P]-5′ triphosphate and T4 polynucleotide kinase). This level of specific activity allows small amounts of DNA to be measured either by autoradiography of gels or filters exposed to film or by direct counting of samples in scintillation fluid.

2) Chemiluminescent detection.

For chemiluminescent detection, digoxigenin-labelled oligonucleotides (Example 1) can be detected using th

chemiluminescent detection system "SOUTHERN LIGHTS," developed by Tropix, Inc. The detection system is diagrammed in Figures 11A and 11B. The technique can be applied to detect DNA that has been captured on either beads, filters, or in solution.

Alkaline phosphatase is coupled to the captured DNA without interfering with the capture system. To do this several methods, derived from commonly used ELISA (Harlow et al.; Pierce, Rockford IL) techniques, can be employed. For example, an antigenic moiety is incorporated into the DNA at sites that will not interfere with (i) the DNA:protein interaction, (ii) the DNA:drug interaction, or (iii) the capture system. In the UL9 DNA:protein/biotin system the DNA has been end-labelled with digoxigenin-11-dUTP (dig-dUTP) and terminal transferase (Example 1, Figure 4). After the DNA was captured and removed from the DNA:protein mixture, an anti-digoxigenin-alkaline phosphatase conjugated antibody was then reacted (Boehringer Mannheim, Indianapolis IN) with the digoxigenin-containing oligonucleotide. The antigenic digoxigenin moiety was recognized by the antibody-enzyme conjugate. The presence of dig-dUTP altered neither the ability of *UL9-COOH* protein to bind the *oriS* SEQ ID NO:1-containing DNA nor the ability of streptavidin to bind the incorporated biotin.

Captured DNA was detected using the alkaline phosphatase-conjugated antibodies to digoxigenin as follows. One chemiluminescent substrate for alkaline phosphatase is 3-(2'-spiroadamantane)-4-methoxy-4-(3"-phosphoryloxy) phenyl-1,2-dioxetane disodium salt (AMPPD) (Example 7). Dephosphorylation of AMPPD results in an unstable compound, which decomposes, releasing a prolonged, steady emission of light at 477 nm. Light measur ment is very sensitive and can detect minute quantities of DNA

(e.g., 10^2-10^3 attomoles) (Example 7).

Colorimetric substrates for the alkaline phosphatase system have also been tested and are useable in the present assay system.

An alternative to the above biotin capture system is to use digoxigenin in place of biotin to modify the oligonucleotide at a site protected by the DNA-binding protein at the assay site: biotin is then used to replace the digoxigenin moieties in the above described detection system. In this arrangement the anti-digoxigenin antibody is used to capture the oligonucleotide probe when it is free of bound protein. Streptavidin conjugated to alkaline phosphatase is then used to detect the presence of captured oligonucleotides.

D) Alternative methods for detecting molecules that increase the affinity of the DNA-binding protein for its cognate site.

In addition to identifying molecules or compounds that cause a decreased affinity of the DNA-binding protein for the screening sequence, molecules may be identified that increase the affinity of the protein for its cognate binding site. In this case, leaving the capture system for unbound DNA in contact with the assay for increasing amounts of time allows the establishment of a fixed off-rate for the DNA:protein interaction (for example SEQ ID NO:1/UL9). In the presence of a stabilizing molecule, the off-rate, as detected by the capture system time points, will be decreased.

Using the capture system for DNA:protein complexes to detect molecules that increase the affinity of the DNA-binding protein for the screening sequence requires that an excess of unlabeled oligonucleotide containing the UL9 binding sit (but not the test sequences) is added to th assay mixture. This is, in effect, an off-rat experiment.

In this case, the control sample (no test molecules or mixtures added) will show a fixed off-rate (ie., samples would be taken at fixed intervals after the addition of the unlabeled competition DNA molecule, applied to nitrocellulose, and a decreasing amount of radiolabeled DNA:protein complex would be observed). In the presence of a DNA-binding test molecule that enhanced the binding of UL9, the off-rate would be decreased (ie., the amount of radiolabeled DNA:protein complexes observed would not decrease as rapidly at the fixed time points as in the control sample).

III. Utility

A. The Usefulness of Sequence-Specific DNA-Binding Molecules.

The present invention defines a high through-put *in vitro* screening assay to test large libraries of biological or chemical mixtures for the presence of DNA-binding molecules having sequence binding preference. The assay is also capable of determining the sequence-specificity and relative affinity of known DNA-binding molecules or purified unknown DNA-binding molecules. Sequence-specific DNA-binding molecules are of particular interest for several reasons, which are listed here. These reasons, in part, outline the rationale for determining the usefulness of DNA-binding molecules as therapeutic agents:

1) Generally, for a given DNA:protein interaction, there are several thousands fewer target DNA-binding sequences per cell than protein molecules that bind to the DNA. Accordingly, even fairly toxic molecules might be delivered in sufficiently low concentration to exert a biological effect by binding to the target DNA sequences.

2) DNA has a r latively mor well-defined structure compared to RNA or protein. Since th general structure of DNA has l ss tertiary structural variation, identifying or

designing specific binding molecules should be easier for DNA than for either RNA or protein. Double-stranded DNA is a repeating structure of deoxyribonucleotides that stack atop one another to form a linear helical structure. In this manner, DNA has a regularly repeating "lattice" structure that makes it particularly amenable to molecular modeling refinements and hence, drug design and development.

3) Many "single-copy" genes (of which there are only 1 or 2 copies in the cell) are transcribed into multiple, potentially thousands, of RNA molecules, each of which may be translated into many proteins. Accordingly, targeting any DNA site, whether it is a regulatory sequence or a coding or noncoding sequence, may require a much lower drug dose than targeting RNAs or proteins.

Proteins (e.g., enzymes, receptors, or structural proteins) are currently the targets of most therapeutic agents. More recently, RNA molecules have become the targets for antisense or ribozyme therapeutic molecules.

4) Blocking the function of a RNA, which encodes a protein, or of a corresponding protein, when that protein regulates several cellular genes, may have detrimental effects: particularly if some of the regulated genes are important for the survival of the cell. However, blocking a DNA-binding site that is specific to a single gene regulated by such a protein results in reduced toxicity.

An example situation (4) is HNF-1 binding to Hepatitis B virus (HBV): HNF-1 binds an HBV enhancer sequence and stimulates transcription of HBV genes (Chang et al.). In a normal cell HNF-1 is a nuclear protein that appears to be important for the regulation of many genes, particularly liver-specific genes (Courtois et al.). If molecules were isolated that specifically bound to the DNA-binding domain of HNF-1, all of the genes regulated by HNF-1 would b down-regulated, including both viral and cellular genes.

Such a drug could be lethal since many of the genes regulated by HNF-1 may be necessary for liver function. However, the assay of the present invention presents the ability to screen for a molecule that could distinguish the HNF-1 binding region of the Hepatitis B virus DNA from cellular HNF-1 sites by, for example, including divergent flanking sequences when screening for the molecule. Such a molecule would specifically block HBV expression without effecting cellular gene expression.

B. General Applications of the Assay.

General applications of the assay include but are not limited to screening libraries of uncharacterized compounds (e.g., biological, chemical or synthetic compounds) for sequence-specific DNA-binding molecules (part III.B.1); determining the sequence-specificity or preference and/or relative affinities of DNA-binding molecules (part III.B.2); and testing of modified derivatives of DNA-binding molecules for altered specificity or affinity (part III.B.3). In particular, since each test compound is screened against up to 4^N sequences, where N is the number of basepairs in the test sequence, the method will generate up to 4^N structure/activity data points for analysing the relationship between compound structure and binding activity, as evidenced by protein binding to an adjacent sequence.

1) Mass-screening of libraries for the presence of sequence-specific DNA-binding molecules.

Many organizations (eg., the National Institutes of Health, pharmaceutical and chemical corporations) have large libraries of chemical or biological compounds from synthetic processes, fermentation broths or extracts that may contain as yet unidentified DNA-binding molecules. One utility of the assay of the present inv ntion is to apply the assay system to the mass-screening of these libraries

of different broths, extracts, or mixtures to detect the specific samples that contain the DNA-binding molecules. Once the specific mixtures that contain the DNA-binding molecules have been identified, the assay has a further usefulness in aiding in the purification of the DNA-binding molecule from the crude mixture.

As purificaiton schemes are applied to the mixture, the assay can be used to test the fractions for DNA-binding activity. The assay is amenable to high throughput (eg., a 96-well plate format automated on robotics equipment such as a Beckman Biomek workstation [Beckman, Palo Alto, CA] with detection using semiautomated plate-reading densitometers, luminometers, or phosphoimagers).

2) The assay of the present invention is also useful for screening molecules that are currently described in the literature as DNA-binding molecules but which have uncertain DNA-binding sequence specificity (ie., having either no well-defined preference for binding to specific DNA sequences or having certain higher affinity binding sites but without defining the relative preference for all possible DNA binding sequences). The assay can be used to determine the specific binding sites for DNA-binding molecules, among all possible choices of sequence that bind with high, low, or moderate affinity to the DNA-binding molecule. Actinomycin D, Distamycin A, and Doxorubicin (Example 6) all provide examples of molecules with these modes of binding. Many anti-cancer drugs, such as Doxorubicin (see Example 6) show binding preference for certain identified DNA sequences, although the absolute highest and lowest specificity sequences have yet to be determined, because, until the invention described herein, the methods (Salas, X. and Portugal, J.; Cullinane, C. and Phillips, D.R.; Phillips, D.R.,; and Phillips, D.R. et al.) f r d tecting differential affinity DNA-binding sites f r any drug were limited. Doxorubicin is one of th most

widely used anti-cancer drugs currently available. As shown in Example 6, Doxorubicin is known to bind some sequences preferentially. Another example of such sequence binding preference is Daunorubicin (Chen et al.) that differs slightly in structure from Doxorubicin (Goodman et al.). Both Daunorubicin and Doxorubicin are members of the anthracycline antibiotic family: antibiotics in this family, and their derivatives, are important antitumor agents (Goodman et al.).

The assay of the present invention allows the sequence preferences or specificities of DNA-binding molecules to be determined. The DNA-binding molecules for which sequence preference or specificity can be determined may include small molecules such as aminoacridines and polycyclic hydrocarbons, planar dyes, various DNA-binding antibiotics and anticancer drugs, as well as DNA-binding macromolecules such as peptides and polymers that bind to nucleic acids (eg, DNA and the derivatized homologs of DNA that bind to the DNA helix).

The molecules that can be tested in the assay for sequence preference/specificity and relative affinity to different DNA sites include both major and minor groove binders as well as intercalating and non-intercalating DNA binders.

3) The assay of the present invention facilitates the identification of different binding activities by molecules derived from known DNA-binding molecules. An example would be to identify derivatives and test these derivatives for DNA-binding activity using the assay of the present invention. Derivatives having DNA-binding activity are then tested for anti-cancer activity through, for example, a battery of assays performed by the National Cancer Institut (Beth sda MD). Further, the assay of the present invention can be used to test derivatives of known anti-canc r ag nts to examin the effect of the modifications

(such as methylation, ethylation and other derivatizations) on DNA-binding activity and specificity. The assay provides (i) an initial screen for the design of better therapeutic derivatives of known agents and (ii) a method to provide a better understanding of the mode of action of such therapeutic derivatives.

4) The screening capacity of this assay is much greater than screening each separate DNA sequence with an individual cognate DNA-binding protein. While direct competition assays involving individual receptor:ligand complexes (eg., a specific DNA:protein complex) are most commonly used for mass screening efforts, each assay requires the identification, isolation, purification, and production of the assay components. Using the assay of the present invention, libraries of synthetic chemicals or biological molecules can be screened for detecting molecules that have preferential binding to virtually any specified DNA sequence using a single assay system. Secondary screens involving the specific DNA:protein interaction may not be necessary, since inhibitory molecules detected in the assay may be tested directly on a biological system (eg., the ability to disrupt viral replication in a tissue culture or animal model).

C. Sequences Targeted by the Assay.

The DNA:protein assay of the present invention has been designed to screen for compounds that bind a full range of DNA sequences that vary in length as well as complexity. Sequence-specific DNA-binding molecules discovered by the assay have potential usefulness as either molecular reagents, therapeutics, or therapeutic precursors. Table I lists several potential specific test sequences. Sequence-specific DNA-binding molecules are potentially powerful therapeutics for essentially any disease or condition that in some way involves DNA.

Examples of test sequences for the assay include: a) binding sequences of factors involved in the maintenance or propagation of infectious agents, especially viruses, bacteria, yeast and other fungi, b) sequences causing the inappropriate expression of certain cellular genes, and c) sequences involved in the replication of rapidly growing cells.

Furthermore, gene expression or replication does not necessarily need to be disrupted by blocking the binding of specific proteins. Specific sequences within coding regions of genes (e.g., oncogenes) are equally valid test sequences since the binding of small molecules to these sequences is likely to perturb the transcription and/or replication of the region. Finally, any molecules that bind DNA with some sequence specificity, that is, not just to one particular test sequence, may be still be useful as anti-cancer agents. Several small molecules with some sequence preference are already in use as anticancer therapeutics. Molecules identified by the present assay may be particularly valuable as lead compounds for the development of congeners (i.e., chemical derivatives of a molecule having differenct specificities) with either different specificity or different affinity.

One advantage of the present invention is that the assay is capable of screening for binding activity directed against any DNA sequence. Such sequences can be medically significant target sequences (see part 1, Medically Significant Target Sites, in this section), scrambled or randomly generated DNA sequences, or well-defined, ordered sets of DNA sequences (see part 2, Ordered Sets of Test Sequences, in this section), which could be used for screening for molecules demonstrating sequence preferential binding (like Doxorubicin) to determine the sequ nces with high st binding affinity and/or to determin the relative relative affinities betw en a large number of different

sequences. There is usefulness in taking either approach for detecting and/or designing new therapeutic agents. Part 3 of this section, Theoretical Considerations for Choosing Target Sequences, outlines the theoretical considerations for choosing DNA target sites in a biological system.

1) Medically significant target sequences.

Few effective viral therapeutics are currently available; yet several potential target sequences for antiviral DNA-binding drugs have been well-characterized. Furthermore, with the accumulation of sequence data on all biological systems, including viral genomes, cellular genomes, pathogen genomes (bacteria, fungi, eukaryotic parasites, etc.), the number of target sites for DNA-binding drugs will increase greatly in the future. Medically significant target sites can be defined as short DNA sequences (approximately 4-30 base pairs) that are required for the expression replication of genetic material. For example, sequences that bind regulatory factors, either transcriptional or replicatory factors, would be ideal target sites for altering gene or viral expression. Secondly, coding sequences may be adequate target sites for disrupting gene function. Thirdly, even non-coding, non-regulatory sequences may be of interest as target sites (e.g., for disrupting replication processes or introducing an increased mutational frequency. Some specific examples of medically significant target sites are shown in Table 1.

TABLE I. MEDICALLY SIGNIFICANT DNA-BINDING SEQUENCES

Test sequence	DNA-binding Protein	Medical Significance
EBV origin of replication	EBNA	infectious mononucleosis, nasal pharyngeal carcinoma
HSV origin of replication	UL9	oral and genital Herpes

VZV origin of replication	UL9-like	shingles
HPV origin of replication	E2	genital warts, cervical carcinoma
Interleukin 2 enhancer	NFAT-1	immunosuppressant
HIV LTR	NFAT-1 NFkB	AIDS, ARC
HBV enhancer	HNF-1	hepatitis
Fibrogen promoter	HNF-1	cardiovascular disease
Oncogene promoter and coding sequences	??	cancer

(Abbreviations: EBV, Epstein-Barr virus; EBNA, Epstein-Barr virus nuclear antigen; HSV, Herpes Simplex virus; VZV, Vericella zoster virus; HPV, human papilloma virus; HIV LTR, Human immunodeficiency virus long terminal repeat; NFAT, nuclear factor of activated T cells; NFkB, nuclear factor kappaB; AIDS, acquired immune deficiency syndrome; ARC, AIDS related complex; HBV, hepatitis B virus; HNF, hepatic nuclear factor.)

The origin of replication binding proteins, Epstein Barr virus nuclear antigen 1 (EBNA-1) (Ambinder, R.F., et al.; Reisman, D. et al.), E2 (which is encoded by the human papilloma virus) (Chin, M.T., et al.), UL9 (which is encoded by herpes simplex virus type 1) (McGeoch,D.J., et al.), and the homologous protein in vericella zoster virus (VZV) (Stow, N.D. and Davison, A.J.), have short, well-defined binding sites within the viral genome and are therefore excellent target sites for a competitive DNA-binding drug. Similarly, recognition sequences for DNA-binding proteins that act as transcriptional regulatory factors are also good target sites for antiviral DNA-binding drugs. Examples include the binding site for h patic nuclear factor (HNF-1), which is r quired for the expression of human hepatitis B virus (HBV) (Chang, H.-K.),

and NFκB and NFAT-1 binding sites in the human immunodeficiency virus (HIV) long terminal repeat (LTR),f one or both of which may be involved in the expression of the virus (Greene, W.C.).

Examples of non-viral DNA targets for DNA-binding drugs are also shown in Table 1 to illustrate the wide range of potential applications for sequence-specific DNA-binding molecules. For example, nuclear factor of activated T cells (NFAT-1) is a regulatory factor that is crucial to the inducible expression of the interleukin 2 (IL-2) gene in response to signals from the antigen receptor, which, in turn, is required for the cascade of molecular events during T cell activation (for review, see Edwards, C.A. and Crabtree, G.R.). The mechanism of action of two immunosuppressants, cyclosporin A and FK506, is thought to be to block the inducible expression of NFAT-1 (Schmidt,A. et al. and Banerji, S.S. et al.). However, the effects of these drugs are not specific to NFAT-1; therefore, a drug targeted specifically to the NFAT-1 binding site in the IL-2 enhancer would be desirable as an improved immunosuppressant.

Targeting the DNA site with a DNA-binding drug rather than targeting with a drug that affects the DNA-binding protein (presumably the target of the current immunosuppressants) is advantageous for at least two reasons: first, there are many fewer target sites for specific DNA sequences than specific proteins (eg., in the case of glucocorticoid receptor, a handful of DNA-binding sites vs. about 50,000 protein molecules in each cell) and secondly, only the targeted gene need be affected by a DNA-binding drug, while a protein-binding drug would disable all the cellular functions of the protein.

An example of the latter point is the binding site f r HNF-1 in the human fibrinogen promoter. Fibrinogen level is one of the most highly correlat d factor with

cardiovascular disease. A drug targeted to either HNF-1 or the HNF-1 binding site in the fibrinogen promoter might be used to decrease fibrinogen expression in individuals at high risk for disease because of the overexpression of fibrinogen. However, since HNF-1 is required for the expression of a number of normal hepatic genes, blocking the HNF-1 protein would be toxic to liver function. In contrast, by blocking a DNA sequence that is composed in part of the HNF-1 binding site and in part by flanking sequences for divergence, the fibrinogen gene can be targeted with a high level of selectivity, without harm to normal cellular HNF-1 functions.

The assay has been designed to screen virtually any DNA sequence. As described above, test sequences of medical significance include viral or microbial pathogen genomic sequences and sequences within or regulating the expression of oncogenes or other inappropriately expressed cellular genes. In addition to the detection of potential antiviral drugs, the assay of the present invention is also applicable to the detection of potential drugs for (i) disrupting the metabolism of other infectious agents, (ii) blocking or reducing the transcription of inappropriately expressed cellular genes (such as oncogenes or genes associated with certain genetic disorders), and (iii) the enhancement or alteration of expression of certain cellular genes.

2) Defined sets of test sequences.

The approach described in the above section discusses screening large numbers of fermentation broths, extracts, or other mixtures of unknowns against specific medically significant DNA target sequences. The assay can also be utilized to screen a large number of DNA sequences against known DNA-binding drugs to determine the relative affinity of th single drug for every possible defined specific

sequence. For example, there are 4^n possible sequences, where n = the number of nucleotides in the sequence. Thus, there are $4^3 = 64$ different three base pair sequences, $4^4 = 256$ different four base pair sequences, $4^5 = 1024$ different 5 base pair sequences, etc. If these sequences are placed in the test site, the site adjacent to the screening sequence (the example used in this invention is the UL9 binding site), then each of the different test sequences can be screened against many different DNA-binding molecules. The test sequences may be placed on either or both sides of the screening sequence, and the sequences flanking the other side of the test sequences are fixed sequences to stabilize the duplex and, on the 3′ end of the top strand, to act as an annealing site for the primer (see Example 1). For example, oligonucleotides sequences could be constructed as shown in Figure 15 (SEQ ID NO:18). In Figure 15 the TEST and SCREENING sequences are indicated.

The preparation of such double-stranded oligonucleotides is described in Example 1 and illustrated in Figure 4A and 4B. The test sequences, denoted in Figure 15 as X:Y (where X = A,C,G, or T and Y = the complementary sequence, T,G,C, or A), may be any of the 256 different 4 base pair sequences shown in Figure 13.

Once a set of test oligonucleotides containing all possible four base pair sequences has been synthesized (see Example 1), the set can be screened with any DNA-binding drug. The relative effect of the drug on each oligonucleotide assay system will reflect the relative affinity of the drug for the test sequence. The entire spectrum of affinities for each particular DNA sequence can therefore be defined for any particular DNA-binding drug. The data generated using this approach can be used to facilitate molecular modeling programs and/or be used directly to design new DNA-binding molecules with incr ased

affinity and specificity.

Another type of ordered set of oligonucleotides that may be useful for screening are sets comprised of scrambled sequences with fixed base composition. For example, if the recognition sequence for a protein is 5′-GATC-3′ and libraries were to be screened for DNA-binding molecules that recognised this sequence, then it would be desirable to screen sequences of similar size and base composition as control sequences for the assay. The most precise experiment is one in which all possible 4 bp sequences are screened; this represents $4^4 = 256$ different test sequences, a number that may not be practical in every situation. However, there are many fewer possible 4 bp sequences with the same base composition (using the bases 1G, 1A, 1T, 1C; n! = 24 different 4 bp sequences with this particular base composition), which provides excellent controls without having to screen large numbers of sequences.

3) Theoretical considerations in choosing biological target sites: Specificity and Toxicity.

One consideration in choosing sequences to screen using the assay of the present invention is test sequence accessibility, that is, the potential exposure of the sequence *in vivo* to binding molecules. Cellular DNA is packaged in chromatin, rendering most sequences relatively inaccessible. Sequences that are actively transcribed, particularly those sequences that are regulatory in nature, are less protected and more accessible to both proteins and small molecules. This observation is substantiated by a large literature on DNAase I sensitivity, footprinting studies with nucleases and small molecules, and general studies on chromatin structure (Tullius). The relative accessibility of a regulatory sequenc , as determined by DNAas I hypersensitivity, is likely to be several ord rs

of magnitude greater than an inactive portion of the cellular genome. For this reason the regulatory sequences of cellular genes, as well as viral regulatory or replication sequences, are useful regions to choose for selecting specific inhibitory small molecules using the assay of the present invention.

Another consideration in choosing sequences to be screened using the assay of the present invention is the uniqueness of the potential test sequence. As discussed above for the nuclear protein HNF-1, it is desirable that small inhibitory molecules are specific to their target with minimal cross reactivity. Both sequence composition and length effect sequence uniqueness. Further, certain sequences are found less frequently in the human genome than in the genomes of other organisms, for example, mammalian viruses. Because of base composition and codon utilization differences, viral sequences are distinctly different from mammalian sequences. As one example, the dinucleotide CG is found much less frequently in mammalian cells than the dinucleotide sequence GC: further, in SV40, a mammalian virus, the sequences AGCT and ACGT are represented 34 and 0 times, respectively. Specific viral regulatory sequences can be chosen as test sequences keeping this bias in mind. Small inhibitory molecules identified which bind to such test sequences will be less likely to interfere with cellular functions.

There are approximately 3×10^9 base pairs (bp) in the human genome. Of the known DNA-binding drugs for which there is crystallographic data, most bind 2-5 bp sequences. There are $4^4 = 256$ different 4 base sequences; therefore, on average, a single 4 bp site is found roughly 1.2×10^7 times in the human genome. An individual 8 base site would be found, on average, about 50,000 times in the genome. On the surface, it might appear that drugs targeted at even an

8 bp site might be deleterious to the cell because there are so many binding sites; however, several other considerations must be recognized. First, most DNA is tightly wrapped in chromosomal proteins and is relatively inaccessible to incoming DNA-binding molecules as demonstrated by the nonspecific endonucleolytic digestion of chromatin in the nucleus (Edwards, C.A. and Firtel, R.A.).

Active transcription units are more accessible than DNA bound in chromosomal proteins, but the most highly exposed regions of DNA in chromatin are the sites that bind regulatory factors. As demonstrated by DNAase I hypersensitivity (Gross, D.S. and Garrard, W.T.), regulatory sites may be 100-1000 times more sensitive to endonucleolytic attack than the bulk of chromatin. This is one reason for targeting regulatory sequences with DNA-binding drugs. Secondly, the argument that several anticancer drugs that bind 2, 3, or 4 bp sequences have sufficiently low toxicity that they can be used as drugs indicates that, if high affinity binding sites for known drugs can be matched with specific viral target sequences, it may be possible to use currently available drugs as antiviral agents at lower concentrations than they are currently used, with a concomitantly lower toxicity.

D. Using Test Matrices and Pattern Matching for the Analysis of Data.

The assay described herein has been designed to use a single DNA:protein interaction to screen for sequence-specific and sequence-preferential DNA-binding molecules that can recognize virtually any specified sequence. By using sequences flanking the recognition site for a single DNA:prot in interaction, a very large numb r of different s quences can be tested. The analysis of data yielded by such exp riments displayed as matrices and analyz d by

pattern matching techniques should yield information about the relatedness of DNA sequences.

The basic principle behind the DNA:protein assay of the present invention is that when molecules bind DNA sequences flanking the recognition sequence for a specific protein the binding of that protein is blocked. Interference with protein binding likely occurs by either (or both) of two mechanisms: 1) directly by steric hindrance, or 2) indirectly by perturbations transmitted to the recognition sequence through the DNA molecule, a type of allosteric perturbation.

Both of these mechanisms will presumably exhibit distance effects. For inhibition by direct steric hindrance direct data for very small molecules is available from methylation and ethylation interference studies. These data suggest that for methyl and ethyl moieties, the steric effect is limited by distance effects to 4-5 base pairs. Even still the number of different sequences that can theoretically be tested for these very small molecules is still very large (i.e., 5 base pair combinations total 4^5 (=1024) different sequences).

In practice, the size of sequences tested can be explored empirically for different sized test DNA-binding molecules. A wide array of sequences with increasing sequence complexity can be routinely investigated. This may be accomplished efficiently by synthesizing degenerate oligonucleotides and multiplexing oligonucleotides in the assay process (i.e., using a group of different oligonucleotides in a single assay) or by employing pooled sequences in test matrices.

In view of the above, assays employing a specific protein and oligonucleotides containing the specific recognition site for that protein flanked by different sequences on either side of th recogniti n site can be

used to simultaneously screen for many different molecules, including small molecules, that have binding preferences for individual sequences or families of related sequences. Figure 12 demonstrates how the analysis of a test matrix yields information about the nature of competitor sequence specificity. As an example, to screen for molecules that could preferentially recognize each of the 256 possible tetranucleotide sequences (Figure 13), oligonucleotides could be constructed that contain these 256 sequences immediately adjacent to a 11 bp recognition sequence of UL9 *oriS* (SEQ ID NO:15), which is identical in each construct.

In Figure 12 "+" indicates that the mixture retards or blocks the formation of DNA:protein complexes in solution and "-" indicates that the mixture had no marked effect on DNA:protein interactions. A summary of the results of the test from Figure 12 are shown in Table .

TABLE 2

Test Mix	Specificity
#1,4,7: oligos	none detected for the above
#2: for recognition site	either nonspecific or specific
#3	AGCT
#5	CATT or ATT
#6	GCATTC, GCATT, CATTC, GCAT, or ATTC
#8	CTTT

These results demonstrate how such a matrix provides data on the presence of sequence specific binding activity is a test mixture and also provides inherent controls for non-specific binding. For example, th effect of test mix #8 on the different test assays reveals that the test mix preferentially affects the oligonucle tides that contain

the sequence CCCT. Note that the sequence does not have to be within the test site for test mix #8 to exert an affect. By displaying the data in a matrix, the analysis of the sequences affected by the different test mixtures is facilitated.

E) Other Applications.

The potential pharmaceutical applications for sequence-specific DNA-binding molecules are broad, including antiviral, antifungal, antibacterial, antitumor agents, immunosuppressants, and cardiovascular drugs. Sequence-specific DNA-binding molecules can also be useful as molecular reagents as, for example, specific sequence probes.

As more molecules are detected, information about the nature of DNA-binding molecules will be gathered, eventually facilitating the design and/or modification of new molecules with different or specialized activities.

Although the assay has been described in terms of the detection of sequence-specific DNA-binding molecules, the reverse assay could be achieved by adding DNA in excess to protein to look for peptide sequence specific protein-binding inhibitors.

The following examples illustrate, but in no way are intended to limit the present invention.

Materials and Methods

Synthetic oligonucleotides were prepared using commercially available automated oligonucleotide synthesizers. Alternatively, custom designed synthetic oligonucleotides may be purchased, for example, from Synthetic Genetics (San Diego, CA). Complementary strands were annealed to generate double-strand oligonucleotides.

Restriction enzymes were obtained fr m Bo hringer Mannheim (Indianapolis IN) or New England Bi labs (Beverly

MA) and were used as per the manufacturer's directions.

Distamycin A and Doxorubicin were obtained from Sigma (St. Louis, MO). Actinomycin D was obtained from Boehringer Mannheim or Sigma.

Example 1

Preparation of the Oligonucleotide Containing the Screening sequence

This example describes the preparation of (i) biotinylated/digoxyginin/radiolabelled, and (ii) radiolabelled double-stranded oligonucleotides that contain the screening sequence and selected Test sequences.

A. Biotinylation.

The oligonucleotides were prepared as described above. The wild-type control sequence for the UL9 binding site, as obtained from HSV, is shown in Figure 4. The screening sequence, i.e. the UL9 binding sequence, is CGTTCGCACTT (SEQ ID NO:1) and is underlined in Figure 4A. Typically, sequences 5′ and/or 3′ to the screening sequence were replaced by a selected Test sequence (Figure 5).

One example of the preparation of a site-specifically biotinylated oligonucleotide is outlined in Figure 4. An oligonucleotide primer complementary to the 3′ sequences of the screening sequence-containing oligonucleotide was synthesized. This oligonucleotide terminated at th residue corresponding to the C in position 9 of th screening sequence. The primer oligonucleotide was hybridized to the oligonucleotide containing the screening sequence. Biotin-11-dUTP (Bethesda Research Laboratories (BRL), Gaithersburg MD) and Klenow enzyme were added t this complex (Figure 4) and the resulting partially double-stranded biotinylated complexes were separated from the unincorporated nucleotides using ith r pre-prepared G-25 Sephadex spin columns (Pharmacia, Piscataway NJ) or "NENSORB" columns (New England Nuclear) as per

manufacturer's instructions. The remaining single-strand region was converted to double-strands using DNA polymerase I Klenow fragment and dNTPs resulting in a fully double-stranded oligonucleotide. A second G-25 Sephadex column was used to purify the double-stranded oligonucleotide. Oligonucleotides were diluted or resuspended in 10 mM Tris-HCl, pH 7.5, 50 mM NaCl, and 1 mM EDTA and stored at -20°C. For radiolabelling the complexes, ^{32}P-alpha-dCTP (New England Nuclear, Wilmington, DE) replaced dCTP for the double-strand completion step. Alternatively, the top strand, the primer, or the fully double-stranded oligonucleotide have been radiolabeled with γ-^{32}P-ATP and polynucleotide kinase (NEB, Beverly, MA). Preliminary studies have employed radiolabeled, double-stranded oligonucleotides. The oligonucleotides are prepared by radiolabeling the primer with T4 polynucleotide kinase and γ-^{32}P-ATP, annealing the "top" strand full length oligonucleotide, and "filling-in" with Klenow fragment and deoxynucleotide triphosphates. After phosphorylation and second strand synthesis, oligonucleotides are separated from buffer and unincorporated triphosphates using G-25 Sephadex preformed spin columns (IBI or Biorad). This process is outlined in Figure 4B. The reaction conditions for all of the above Klenow reactions were as follows: 10 mM Tris-HCl, pH 7.5, 10 mM $MgCl_2$, 50 mM NaCl, 1 mM dithioerythritol, 0.33-100 μM deoxytriphosphates, 2 units Klenow enzyme (Boehringer-Mannheim, Indianapolis IN). The Klenow reactions were incubated at 25°C for 15 minutes to 1 hour. The polynucleotide kinase reactions were incubated at 37°C for 30 minutes to 1 hour.

B) End-labeling with digoxigenin. The biotinylated, radiolabelled oligonucleotides or radiolabeled oligonucl otides were isolated as above and resuspended in 0.2 M potassium cacodylate (pH=7.2), 4 mM $MgCl_2$, 1 mM 2-

mercaptoethanol, and 0.5 mg/ml bovine serum albumin. To this reaction mixture digoxigenin-11-dUTP (an analog of dTTP, 2′-deoxy-uridine-5′-triphosphate, coupled to digoxigenin via an 11-atom spacer arm, Boehringer Mannheim, Indianapolis IN) and terminal deoxynucleotidyl transferas (GIBCO BRL, Gaithersburg, MD) were added. The number of Dig-11-dUTP moieties incorporated using this method appeared to be less than 5 (probably only 1 or 2) as judged by electrophoretic mobility on polyacrylamide gels of the treated fragment as compared to oligonucleotides of known length.

The biotinylated or non-biotinylated, digoxygenin-containing, radiolabelled oligonucleotides were isolated as above and resuspended in 10 mM Tris-HCl, 1 mM EDTA, 50 mM NaCl, pH 7.5 for use in the binding assays.

The above procedure can also be used to biotinylate the other strand by using an oligonucleotide containing the screening sequence complementary to the one shown in Figure 4 and a primer complementary to the 3′ end of that molecule. To accomplish the biotinylation Biotin-7-dATP was substituted for Biotin-11-dUTP. Biotinylation was also accomplished by chemical synthetic methods: for example, an activated nucleotide is incorporated into the oligonucleotide and the active group is subsequently reacted with NHS-LC-Biotin (Pierce). Other biotin derivatives can also be used.

C. Radiolabelling the Oligonucleotides

Generally, oligonucleotides were radiolabelled with gamma-^{32}P-ATP or alpha-^{32}P-deoxynucleotide triphosphates and T4 polynucleotide kinase or the Klenow fragment of DNA polymerase, respectively. Labelling reactions were performed in the buffers and by the methods recommended by th manufacturers (New England Biolabs, Beverly MA; Bethesda Research Laboratories, Gaithersburg MD; or

Boehringer/Mannheim, Indianapolis IN). Oligonucleotides were separated from buffer and unincorporated triphosphates using G-25 Sephadex preformed spin columns (IBI, New Haven, CT; or Biorad, Richmond, CA) or "NENSORB" preformed columns (New England Nuclear, Wilmington, DE) as per the manufacturers instructions.

There are several reasons to enzymatically synthesize the second strand. The two main reasons are that by using an excess of primer, second strand synthesis can be driven to near completion so that nearly all top strands are annealed to bottom strands, which prevents the top strand single strands from folding back and creating additional and unrelated double-stranded structures, and secondly, since all of the oligonucleotides are primed with a common primer, the primer can bear the end-label so that all of the oligonucleotides will be labeled to exactly the same specific activity.

Example 2

Preparation of the UL9 Protein

A. Cloning of the UL9 coding sequences into pAC373.

To express full length UL9 protein a baculovirus expression system has been used. The sequence of the UL9 coding region of Herpes Simplex Virus has been disclosed by McGeoch et al. and is available as an EMBL nucleic acid sequence. The recombinant baculovirus AcNPV/UL9A, which contained the UL9 coding sequence, was obtained from Mark Challberg (National Institutes of Health, Bethesda MD). The construction of this vector has been previously described (Olivo et al. (1988, 1989)). Briefly, the *NarI/-EcoRV* fragment was derived from pMC160 (Wu et al.). Blunt-ends were generated on this fragment by using all four dNTPs and the Klenow fragment of DNA polymerase I (Boehringer Mannheim, Indianapolis IN) to fill in the terminal ov rhangs. The resulting fragment was blunt-end ligated into the unique *BamHI* site of the baculoviral

vector pAC3T3 (Summers et al.).

B. Cloning of the UL9 coding sequence in pVL1393

The UL9 coding region was cloned into a second baculovirus vector, pVL1393 (Luckow et al.). The 3077 bp *NarI/EcoRV* fragment containing the UL9 gene was excised from vector pEcoD (obtained from Dr. Bing Lan Rong, Eye Research Institute, Boston, MA): the plasmid pEcoD contains a 16.2 kb EcoRI fragment derived from HSV-I that bears the UL9 gene (Goldin et al.). Blunt-ends were generated on the UL9-containing fragment as described above. *EcoRI* linkers (10 mer) were blunt-end ligated (Ausubel et al.; Sambrook et al.) to the blunt-ended *NarI/-EcoRV* fragment.

The vector pVL1393 (Luckow et al.) was digested with *EcoRI* and the linearized vector isolated. This vector contains 35 nucleotides of the 5′ end of the coding region of the polyhedron gene upstream of the polylinker cloning site. The polyhedron gene ATG has been mutated to ATT to prevent translational initiation in recombinant clones that do not contain a coding sequence with a functional ATG. The *EcoRI*/UL9 fragment was ligated into the linearized vector, the ligation mixture transformed into *E. coli* and ampicillin resistant clones selected. Plasmids recovered from the clones were analyzed by restriction digestion and plasmids carrying the insert with the amino terminal UL9 coding sequences oriented to the 5′ end of the polyhedron gene were selected. This plasmid was designated pVL1393/UL9 (Figure 7).

pVL1393/UL9 was cotransfected with wild-type baculoviral DNA (AcMNPV; Summers et al.) into SF9 (*Spodoptera frugiperda*) cells (Summers et al.). Recombinant baculovirus-infected Sf9 cells were identified and clonally purified (Summers et al.).

C. Expr ssion of the UL9 Protein.

Clonal isolates of recombinant baculovirus infected

Sf9 cells were grown in Grace's medium as described by Summers et al. The cells were scraped from tissue culture plates and collected by centrifugation (2,000 rpm, for 5 minutes, 4°C). The cells were then washed once with phosphate buffered saline (PBS) (Maniatis et al.). Cell pellets were frozen at -70°C. For lysis the cells were resuspended in 1.5 volumes 20 mM HEPES, pH 7.5, 10% glycerol, 1.7 M NaCl, 0.5 mM EDTA, 1 mM dithiothreitol (DTT), and 0.5 mM phenyl methyl sulfonyl fluoride (PMSF). Cell lysates were cleared by ultracentrifugation (Beckman table top ultracentrifuge, TLS 55 rotor, 34 krpm, 1 hr, 4°C). The supernatant was dialyzed overnight at 4°C against 2 liters dialysis buffer (20 mM HEPES, pH 7.5, 10% glycerol, 50 mM NaCl, 0.5 mM EDTA, 1 mM dtt, and 0.1 mM PMSF).

These partially purified extracts were prepared and used in DNA:protein binding experiments. If necessary extracts were concentrated using a "CENTRICON 30" filtration device (Amicon, Danvers MA).

D. Cloning the Truncated UL9 Protein.

The sequence encoding the C-terminal third of UL9 and the 3′ flanking sequences, an approximately 1.2 kb fragment, was subcloned into the bacterial expression vector, pGEX-2T (Figure 6). The pGEX-2T is a modification of the pGEX-1 vector of Smith et al. which involved the insertion of a thrombin cleavage sequence in-frame with the glutathione-S-transferase protein (gst).

A 1,194 bp *BamHI/EcoRV* fragment of pEcoD was isolated that contained a 951 bp region encoding the C-terminal 317 amino acids of UL9 and 243 bp of the 3′ untranslated region.

This *BamHI/EcoRV* UL9 carboxy-terminal (UL9-COOH) containing fragment was blunt-ended and *EcoRI* linkers add d as described above. The *EcoRI* linkers were designed to

allow in-frame fusion of the UL9 coding sequence to the gst-thrombin coding sequences. The linkered fragment was isolated and digested with *EcoRI*. The pGEX-2T vector was digested with *EcoRI*, treated with Calf Intestinal Alkaline Phosphatase (CIP) and the linear vector isolated. The *EcoRI* linkered UL9-COOH fragment was ligated to the linear vector (Figure 6). The ligation mixture was transformed into *E. coli* and ampicillin resistant colonies were selected. Plasmids were isolated from the ampicillin resistant colonies and analyzed by restriction enzyme digestion. A plasmid which generated a gst/thrombin/UL9-COOH in frame fusion was identified (Figure 6) and designated pGEX-2T/UL9-COOH.

A. Expression of the Truncated UL9 Protein.

E. coli strain JM109 was transformed with pGEX-2T/C-UL9-COOH and was grown at 37°C to saturation density overnight. The overnight culture was diluted 1:10 with LB medium containing ampicillin and grown from one hour at 30°C. IPTG (isopropyllthio-β-galactoside) (GIBCO-BRL) was added to a final concentration of 0.1 mM and the incubation was continued for 2-5 hours. Bacterial cells containing the plasmid were subjected to the temperature shift and IPTG conditions, which induced transcription from the *tac* promoter.

Cells were harvested by centrifugation and resuspended in 1/100 culture volume of MTPBS (150 mM NaCl, 16 mM Na_2HPO_4, 4 mM NaH_2PO_4). Cells were lysed by sonication and lysates cleared of cellular debris by centrifugation.

The fusion protein was purified over a glutathione agarose affinity column as described in detail by Smith et al. The fusion protein was eluted from the affinity column with reduced glutathione, dialyzed against UL9 dialysis buffer (20 mM HEPES pH 7.5, 50 mM NaCl, 0.5 mM EDTA, 1 mM DTT, 0.1 mM PMSF) and cleaved with thrombin (2 ng/ug of

fusion protein).

An aliquot of the supernatant obtained from IPTG-induced cultures of pGEX-2T/C-UL9-COOH-containing cells and an aliquot of the affinity-purified, thrombin-cleaved protein were analyzed by SDS-polyacrylamide gel electrophoresis. The result of this analysis is shown in Figure 8. The 63 kilodalton GST/C-UL9 fusion protein is the largest band in the lane marked GST-UL9 (lane 2). The first lane contains protein size standards. The UL9-COOH protein band (lane GST-UL9 + Thrombin, Figure 8, lane 3) is the band located between 30 and 46 kD: the glutathione transferase protein is located just below the 30 kD size standard. In a separate experiment a similar analysis was performed using the uninduced culture: it showed no protein corresponding in size to the fusion protein.

Extracts are dialyzed before use. Also, if necessary, the extracts can be concentrated typically by filtration using a "CENTRICON 30" filter.

Example 3

Binding Assays

A. Band shift gels.

DNA:protein binding reactions containing both labelled complexes and free DNA were separated electrophoretically on 4-10% polyacrylamide/Tris-Borate-EDTA (TBE) gels (Freid et al.; Garner et al.). The gels were then fixed, dried, and exposed to X-ray film. The autoradiograms of the gels were examined for band shift patterns.

B. Filter Binding Assays

A second method used particularly in determining the off-rates for protein:oligonucleotide complexes is filter binding (Woodbury et al.). Nitrocellulose disks (Schleicher and Schuell, BA85 filters) that have been soaked in binding buffer (see below) were placed on a vacuum filter apparatus. DNA:protein binding reactions

(see below; typically 15-30 µl) are diluted to 0.5 ml with binding buffer (this dilutes the concentration of components without dissociating complexes) and applied to the discs with vacuum applied. Under low salt conditions the DNA:protein complex sticks to the filter while free DNA passes through. The discs are placed in scintillation counting fluid (New England Nuclear), and the cpm determined using a scintillation counter.

This technique has been adapted to 96-well and 72-slot nitrocellulose filtration plates (Schleicher and Schuell) using the above protocol except (i) the reaction dilution and wash volumes are reduced and (ii) the flow rate through the filter is controlled by adjusting the vacuum pressure. This method greatly facilitates the number of assay samples that can be analyzed. Using radioactive oligonucleotides, the samples are applied to nitrocellulose filters, the filters are exposed to x-ray film, then analyzed using a Molecular Dynamics scanning densitometer. This system can transfer data directly into analytical software programs (e.g., Excel) for analysis and graphic display.

Example 4

Functional UL9 Binding Assay

A. Functional DNA-binding Activity Assay

Purified protein was tested for functional activity using band-shift assays. Radiolabelled oligonucleotides (prepared as in Example 1B) that contain the 11 bp recognition sequence were mixed with the UL9 protein in binding buffer (optimized reaction conditions: 0.1 ng ^{32}P-DNA, 1 ul UL9 extract, 20 mM HEPES, pH 7.2, 50 mM KCl, and 1 mM DTT). The reactions were incubated at room temperature for 10 minutes (binding occurs in less than 2 minut s), then separated lectrophoretically on 4-10% non-denaturing polyacrylamide gels. UL9-specific binding to

the oligonucleotide is indicated by a shift in mobility of the oligonucleotide on the gel in the presence of the UL9 protein but not in its absence. Bacterial extracts containing (+) or without (-) UL9 protein and affinity purified UL9 protein were tested in the assay. Only bacterial extracts containing UL9 or affinity purified UL9 protein generate the gel band-shift indicating protein binding.

The degree of extract that needed to be added to the reaction mix, in order to obtain UL9 protein excess relative to the oligonucleotide, was empirically determined for each protein preparation/extract. Aliquots of the preparation were added to the reaction mix and treated as above. The quantity of extract at which the majority of the labelled oligonucleotide appears in the DNA:protein complex was evaluated by band-shift or filter binding assays. The assay is most sensitive under conditions in which the minimum amount of protein is added to bind most of the DNA. Excess protein can decrease the sensitivity of the assay.

B. Rate of Dissociation

The rate of dissociation is determined using a competition assay. An oligonucleotide having the sequence presented in Figure 4, which contained the binding site for UL9 (SEQ ID NO:14), was radiolabelled with ^{32}P-ATP and polynucleotide kinase (Bethesda Research Laboratories). The competitor DNA was a 17 base pair oligonucleotide (SEQ ID NO:16) containing the binding site for UL9.

In the competition assays, the binding reactions (Example 4A) were assembled with each of the oligonucleotides and placed on ice. Unlabelled oligonucleotide (1 μg) was added 1, 2, 4, 6, or 21 hours before loading the r action on an 8% polyacrylamide gel (run in TBE buffer (Maniatis et al.)) to separate the

reaction components. The dissociation rates, under these conditions, for the truncated UL9 (UL9-COOH) and the full length UL9 is approximately 4 hours at 4°C. In addition, random oligonucleotides (a 10,000-fold excess) that did not contain the UL9 binding sequence and sheared herring sperm DNA (a 100,000-fold excess) were tested: neither of these control DNAs competed for binding with the oligonucleotide containing the UL9 binding site.

C. Optimization of the UL9 Binding Assay

(i) Truncated UL9 from the bacterial expression system.

The effects of the following components on the binding and dissociation rates of UL9-COOH with its cognate binding site have been tested and optimized: buffering conditions (including the pH, type of buffer, and concentration of buffer); the type and concentration of monovalent cation; the presence of divalent cations and heavy metals; temperature; various polyvalent cations at different concentrations; and different redox reagents at different concentrations. The effect of a given component was evaluated starting with the reaction conditions given above and based on the dissociation reactions described in Example 4B.

The optimized conditions used for the binding of UL9-COOH contained in bacterial extracts (Example 2E) to oligonucleotides containing the HSV ori sequence (SEQ ID NO:1) were as follows: 20 mM HEPES, pH 7.2, 50 mM KCl, 1 mM DTT, 0.005 - 0.1 ng radiolabeled (specific activity, approximately 10^8 cpm/μg) or digoxiginated, biotinylated oligonucleotide probe, and 5-10 μg crude UL9-COOH protein preparation (1 mM EDTA is optional in the reaction mix). Under optimized conditions, UL9-COOH binds very rapidly and has a dissociati n rate of about 4 hours at 4°C with non-biotinylated oligonucleotide and 5-10 minutes with

biotinylated oligonucleotides. The dissociation rate of UL9-COOH changes markedly under different physical conditions. Typically, the activity of a UL9 protein preparation was assessed using the gel band-shift assay and related to the total protein content of the extract as a method of standardization. The addition of herring sperm DNA depended on the purity of UL9 used in the experiment Binding assays were incubated at 25°C for 5-30 minutes.

(ii) Full length UL9 protein from the baculovirus system.

The binding reaction conditions for the full length baculovirus-produced UL9 polypeptide have also been optimized. The optimal conditions for the current assay were determined to be as follows: 20 mM Hepes; 100 mM NaCl; 0.5 mM dithiothreitol; 1 mM EDTA; 5% glycerol; from 0 to 10^4-fold excess of sheared herring sperm DNA; 0.005 - 0.1 ng radiolabeled (specific activity, approximately 10^8 cpm/μg) or digoxiginated, biotinylated oligonucleotide probe, and 5-10 μg crude UL9 protein preparation. The full length protein also binds well under the optimized conditions established for the truncated UL9-COOH protein.

Example 5

The Effect of Test Sequence Variation on the Off-Rate of the UL9 Protein

The oligonucleotides shown in Figure 5 were radiolabelled as described above. The competition assays were performed as described in Example 4B using UL9-COOH. Radiolabelled oligonucleotides were mixed with the UL9-COOH protein in binding buffer (typical reaction: 0.1 ng oligonucleotide ^{32}P-DNA, 1 μl UL9-COOH extract, 20 mM HEPES, pH 7.2, 50 mM KCl, 1 mM EDTA, and 1 mM DTT). The reactions wer incubated at room temperature for 10 minutes. A zero time point sample was then taken and loaded onto an 8%

polyacrylamide gel (run use TBE). One μg of the unlabelled 17 bp competitive DNA oligonucleotide (SEQ ID NO:16) (Example 4B) was added at 5, 10, 15, 20, or 60 minutes before loading the reaction sample on the gel. The results of this analysis are shown in Figure 9: the screening sequences that flank the UL9 binding site (SEQ ID NO:5-SEQ ID NO:13) are very dissimilar but have little effect on the off-rate of UL9. Accordingly, these results show that th UL9 DNA binding protein is effective to bind to a screening sequence in duplex DNA with a binding affinity that is substantially independent of test sequences placed adjacent the screening sequence. Filter binding experiments gave the same result.

Example 6

The Effect of Actinomycin D, Distamycin A, and Doxorubicin on UL9 Binding to the screening Sequence is Dependent on the Specific Test Sequence

Different oligonucleotides, each of which contained the screening sequence (SEQ ID NO:1) flanked on the 5′ and 3′ sides by a test sequence (SEQ ID NO:5 to SEQ ID NO:13), were evaluated for the effects of distamycin A, actinomycin D, and doxorubicin on UL9-COOH binding.

Binding assays were performed as described in Example 5. The oligonucleotides used in the assays are shown in Figure 5. The assay mixture was allowed to pre-equilibrate for 15 minutes at room temperature prior to the addition of drug.

A concentrated solution of Distamycin A was prepared in dH_20 and was added to the binding reactions at the following concentrations: 0, 1 μM, 4 μM, 16 μM, and 40 μM. Th drug was added and incubated at room temperatur for 1 h ur. The reaction mixtures were then loaded n an 8% polyacrylamide gel (Example 5) and the c mponents separated

electrophoretically. Autoradiographs of these gels are shown in Figure 10A. The test sequences tested were as follows: UL9 polyT, SEQ ID NO:9; UL9 CCCG, SEQ ID NO:5; UL9 GGGC, SEQ ID NO:6; UL9 polyA, SEQ ID NO:8; and UL9 ATAT, SEQ ID NO:7. These results demonstrate that Distamycin A preferentially disrupts binding to UL9 polyT, UL9 polyA and UL9 ATAT.

A concentrated solution of Actinomycin D was prepared in dH_20 and was added to the binding reactions at the following concentrations: 0 μM and 50 μM. The drug was added and incubated at room temperature for 1 hour. Equal volumes of dH_20 were added to the control samples. The reaction mixtures were then loaded on an 8% polyacrylamide gel (Example 5) and the components separated electrophoretically. Autoradiographs of these gels are shown in Figure 10B. In addition to the test sequences tested above with Distamycin A, the following test sequences were also tested with Actinomycin D: ATori1, SEQ ID NO:11; oriEco2, SEQ ID NO:12, and oriEco3, SEQ ID NO:13. These results demonstrate that actinomycin D preferentially disrupts the binding of UL9 to the oligonucleotides UL9 CCCG and UL9 GGGC.

A concentrated solution of Doxorubicin was prepared in dH_20 and was added to the binding reactions at the following concentrations: 0 μM, 15 μM and 35 μM. The drug was added and incubated at room temperature for 1 hour. Equal volumes of dH_20 were added to the control samples. The reaction mixtures were then loaded on an 8% polyacrylamide gel (Example 5) and the components separated electrophoretically. Autoradiographs of these gels are shown in Figure 10C. The same test sequences were tested as for Actinomycin D. These results demonstrate that Doxorubicin preferentially disrupts the binding of UL9 to the oligonucleotides UL9polyT, UL9 GGGC, oriEco2, and

oriEco3. Doxorubicin appears to particularly disrupt the UL9:screening sequence interaction when the test sequence oriEco3 is used. The sequences of the test sequences for oriEco2 and oriEco3 differ by only one base: an additional T residue inserted at position 12, compare SEQ ID NO:12 and SEQ ID NO:13.

Example 7

Use of the Biotin/Streptavidin Reporter System

A. The Capture of Protein-Free DNA.

Several methods have been employed to sequester unbound DNA from DNA:protein complexes.

(i) Magnetic beads

Streptavidin-conjugated superparamagnetic polystyrene beads (Dynabeads M-280 Streptavidin, Dynal AS, 6-7x10^8 beads/ml) are washed in binding buffer then used to capture biotinylated oligonucleotides (Example 1). The beads are added to a 15 ul binding reaction mixture containing binding buffer and biotinylated oligonucleotide. The beads/oligonucleotide mixture is incubated for varying lengths of time with the binding mixture to determine the incubation period to maximize capture of protein-free biotinylated oligonucleotides. After capture of the biotinylated oligonucleotide, the beads can be retrieved by placing the reaction tubes in a magnetic rack (96-well plate magnets are available from Dynal). The beads are then washed.

(ii) Agarose beads

Biotinylated agarose beads (immobilized D-biotin, Pierce, Rockford, IL) are bound to avidin by treating the beads with 50 μg/μl avidin in binding buffer overnight at 4°C. The beads ar washed in binding buffer and used to capture biotinylated DNA. The beads ar mixed with binding mixtures to capture biotinylated DNA. The beads are

removed by centrifugation or by collection on a non-binding filter disc.

For either of the above methods, quantification of the presence of the oligonucleotide depends on the method of labelling the oligonucleotide. If the oligonucleotide is radioactively labelled: (i) the beads and supernatant can be loaded onto polyacrylamide gels to separate protein:DNA complexes from the bead:DNA complexes by electrophoresis, and autoradiography performed; (ii) the beads can be placed in scintillation fluid and counted in a scintillation counter. Alternatively, presence of the oligonucleotide can be determined using a chemiluminescent or colorimetric detection system.

B. Detection of Protein-Free DNA.

The DNA is end-labelled with digoxigenin-11-dUTP (Example 1). The antigenic digoxigenin moiety is recognized by an antibody-enzyme conjugate, anti-digoxigenin-alkaline phosphatase (Boehringer Mannheim Indianapolis IN). The DNA/antibody-enzyme conjugate is then exposed to the substrate of choice. The presence of dig-dUTP does not alter the ability of protein to bind the DNA or the ability of streptavidin to bind biotin.

(i) Chemiluminescent Detection.

Digoxigenin-labelled oligonucleotides are detected using the chemiluminescent detection system "SOUTHERN LIGHTS" developed by Tropix, Inc. (Bedford, MA). Use of this detection system is illustrated in Figures 11A and 11B. The technique can be applied to detect DNA that has been captured on either beads or filters.

Biotinylated oligonucleotides, which have terminal digoxygenin-containing residues (Example 1), are captured on magnetic (Figure 11A) r agarose beads (Figure 11B) as described above. The beads are isolated and treated to block non-specific binding by incubation with I-Light

blocking buffer (Tropix) for 30 minutes at room temperature. The presence of oligonucleotides is detected using alkaline phosphatase-conjugated antibodies to digoxygenin. Anti-digoxigenin-alkaline phosphatase (anti-dig-AP, 1:5000 dilution of 0.75 units/ul, Boehringer Mannheim) is incubated with the sample for 30 minutes, decanted, and the sample washed with 100 mM Tris-HCl, pH 7.5, 150 mM NaCl. The sample is pre-equilibrated with 2 washes of 50 mM sodium bicarbonate, pH 9.5, 1 M $MgCl_2$, then incubated in the same buffer containing 0.25 mM 3-(2'-spiroadamantane)-4-methoxy-4-(3'-phosphoryloxy) phenyl-1,2-dioxetane disodium salt (AMPPD) for 5 minutes at room temperature. AMPPD was developed (Tropix Inc.) as a chemiluminescent substrate for alkaline phosphatase. Upon dephosphorylation of AMPPD the resulting compound decomposes, releasing a prolonged, steady emission of light at 477 nm.

Excess liquid is removed from filters and the emission of light occurring as a result of the dephosphorylation of AMPPD by alkaline phosphatase can be measured by exposure to x-ray film or by detection in a luminometer.

In solution, the bead-DNA-anti-dig-AP is resuspended in "SOUTHERN LIGHT" assay buffer and AMPPD and measured directly in a luminometer. Large scale screening assays are performed using a 96-well plate-reading luminometer (Dynatech Laboratories, Chantilly, VA). Subpicogram quantities of DNA (10^2 to 10^3 attomoles (an attomole is 10^{-18} moles)) can be detected using the Tropix system in conjunction with the plate-reading luminometer.

(ii) Colorimetric Detection.

Standard alkalin phosphatase colorimetric substrates are also suitable for the above detection reactions. Typically substrates include 4-nitrophenyl phosphate

(Boehringer Mannheim). Results of colorimetric assays can be evaluated in multiwell plates (as above) using a plate-reading spectrophotometer (Molecular Devices, Menlo Park CA). The use of the light emission system is more sensitive than the colorimetric systems.

While the invention has been described with reference to specific methods and embodiments, it will be appreciated that various modifications and changes may be made without departing from the invention.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Edwards, Cynthia A.
Cantor, Charles R.
Andrews, Beth M.

(ii) TITLE OF INVENTION: Screening Assay for the Detection of DNA-Binding Molecules

(iii) NUMBER OF SEQUENCES: 18

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: LAW OFFICES OF PETER J. DEHLINGER
(B) STREET: P.O. Box 60850
(C) CITY: Palo Alto
(D) STATE: CA
(E) COUNTRY: USA
(F) ZIP: 94306

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:

(vii) PREVIOUS APPLICATION DATA:
(A) APPLICATION NUMBER: 07/723,618
(B) FILING DATE: 27-JUN-1991
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Fabian, Gary R.
(B) REGISTRATION NUMBER: 33,875
(C) REFERENCE/DOCKET NUMBER: 4600-0075.41

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (415) 323-8302
(B) TELEFAX: (415) 323-8306

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(C) INDIVIDUAL ISOLATE: UL9 BINDING SITE, HSV oriS, higher affinity

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

CGTTCGCACT T 11

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(C) INDIVIDUAL ISOLATE: UL9 BINDING SITE, HSV oriS, lower affinity

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

TGCTCGCACT T 11

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(C) INDIVIDUAL ISOLATE: UL9Z1 TEST SEQ. / UL9 ASSAY SEQ.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

GCGCGCGCGC GTTCGCACTT CCGCCGCCGG 30

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(C) INDIVIDUAL ISOLATE: UL9Z2 TEST SEQ. / UL9 ASSAY SEQ.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

GGCGCCGGCC GTTCGCACTT CGCGCGCGCG 30

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(C) INDIVIDUAL ISOLATE: UL9 CCCG TEST SEQ. / UL9 ASSAY SEQ.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

GGCCCGCCCC GTTCGCACTT CCCGCCCCGG 30

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(C) INDIVIDUAL ISOLATE: UL9 GGGC TEST SEQ. / UL9 ASSAY SEQ.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

GGCGGGCGCC GTTCGCACTT GGGCGGGCGG 30

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(C) INDIVIDUAL ISOLATE: UL9 ATAT TEST SEQ. / UL9 ASSAY SEQ.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

GGATATATAC GTTCGCACTT TAATTATTGG 30

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(C) INDIVIDUAL ISOLATE: UL9 polyA TEST SEQ. / UL9 ASSAY SEQ.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

GGAAAAAAAC GTTCGCACTT AAAAAAAAGG 30

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(C) INDIVIDUAL ISOLATE: UL9 polyT TEST SEQ. / UL9 ASSAY SEQ.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

GGTTTTTTTC GTTCGCACTT TTTTTTTTGG 30

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(C) INDIVIDUAL ISOLATE: UL9 GCAC TEST SEQ. / UL9 ASSAY SEQ.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

GGACGCACGC GTTCGCACTT GCAGCAGCGG 30

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(C) INDIVIDUAL ISOLATE: UL9 ATori-1 TEST SEQUENCE / UL9 ASSAY SEQ.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GCGTATATAT CGTTCGCACT TCGTCCCAAT 30

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(C) INDIVIDUAL ISOLATE: oriECO2 TEST SEQ. / UL9 ASSAY SEQ.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

GGCGAATTCG ACGTTCGCAC TTCGTCCCAA T 31

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(C) INDIVIDUAL ISOLATE: oriECO3 TEST SEQ. / UL9 ASSAY SEQ.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

GGCGAATTCG ATCGTTCGCA CTTCGTCCCA AT 32

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 36 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(C) INDIVIDUAL ISOLATE: WILD TYPE

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

AAGTGAGAAT TCGAAGCGTT CGCACTTCGT CCCAAT 36

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: TRUNCATED UL9 BINDING SITE, COMPARE SEQ ID NO:1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

TTCGCACTT 9

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(C) INDIVIDUAL ISOLATE: HSVB1/4, SEQUENCE OF COMPETITOR DNA MOLECULE

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

GGTCGTTCGC ACTTCGC 17

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(C) INDIVIDUAL ISOLATE: UL9 BINDING SITE, HSV oriS

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

CGTTCTCACTT 11

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 37 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(C) INDIVIDUAL ISOLATE: UL9 ASSAY SEQUENCE, FIGURE 15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

GCGTAANNNNCGTTCGCACTTNNNNCTTCGTCCCAAT 37

IT IS CLAIMED:

1. A method of screening for molecules capable of binding to a selected test sequence in a duplex DNA, comprising

(i) adding a molecule to be screened to a test system composed of (a) a DNA binding protein which is effective to bind to a screening sequence in a duplex DNA with a binding affinity that is substantially independent of said test sequence adjacent the screening sequence, but where said protein binding is sensitive to binding of molecules to such test sequence, and (b) a duplex DNA having said screening and test sequences adjacent one another, wherein the binding protein is present in molar excess over the screening sequence present in the duplex DNA,

(ii) incubating the molecule in the test system for a period sufficient to permit binding of the compound being tested to the test sequence in the duplex DNA, and

(iii) detecting the amount of binding protein bound to the duplex DNA before and after said adding.

2. The method of claim 1, wherein the screening sequence/binding protein is selected from the group consisting of EBV origin of replication/EBNA, HSV origin of replication/UL9, VZV origin of replication/UL9-like, and HPV origin of replication/E2, and lambda o_L-o_R/cro.

3. The method of claim 2, wherein the DNA screening sequence is from the HSV origin of replication and the binding protein is UL9.

4. The method of claim 3, wherein the DNA screening sequence is selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:15, and SEQ ID NO:17.

5. The method of claim 1, wherein said detecting is accomplished using either a gel band-shift assay or a filter-binding assay.

6. The method of claim 1, wherein the test sequences are selected from the group consisting of EBV origin of replication, HSV origin of replication, VZV origin of replication, HPV origin of replication, interleukin 2 enhancer, HIV-LTR, HBV enhancer, and fibrinogen promoter.

7. The method of claim 1, where the test sequences are selected from a defined set of nucleic acid sequences.

8. The method of claim 7, wherein said defined set of DNA sequences has $[X_N]^N$ combinations, where X_N is sequence of deoxyribonucleotides and the number of deoxyribonucleotides in each sequence is N, N is greater than or equal to three.

9. The method of claim 8, wherein N is 3-20.

10. The method of claim 9, wherein N is 4-10.

11. The method of claim 10, wherein N is 4 and the number of combinations is 256.

12. The method of claim 1, wherein said detecting includes the use of a capture system that traps DNA free of bound protein.

13. The method of claim 12, wherein the capture system involves the biotinylation of a nucleotide within th screening sequence (i) that does not eliminate th protein's ability to bind to the scr ening sequ nc , (ii)

that is capable of binding streptavidin, and (iii) wherein the biotin moiety is protected from interactions with streptavidin when the protein is bound to the screening sequence.

14. The method of claim 1, wherein said binding protein is present in a molar concentration less than or equal to the molar concentration of the screening sequence present in the duplex DNA.

15. The method of claim 7, wherein said defined set of nucleic acid sequences are all possible sequential combinations of a number of deoxyribonucleotides, N, wherein N is less than 20 and more than 2.

16. The method of claim 15, wherein N is less than 10 and more than 2.

17. The method of claim 16, wherein N is 4.

18. A screening system for identifying molecules that are capable of binding to a test sequence in a target duplex DNA sequence, comprising

a duplex DNA having screening and test sequences adjacent one another,

a DNA binding protein that is effective in binding to said screening sequence in the duplex DNA with a binding affinity that is substantially independent of said test sequence adjacent the screening sequence, but which is sensitive to binding of molecules to said test sequence, wherein the binding protein is present in molar excess over the screening sequence present in the duplex DNA, and means for detecting the amount of binding protein bound to the DNA.

19. The system of claim 18, wherein the test sequences are selected from the group consisting of EBV origin of replication, HSV origin of replication, VZV origin of replication, HPV origin of replication, interleukin 2 enhancer, HIV-LTR, HBV enhancer, and fibrinogen promoter.

20. The system of claim 18, where the test sequences are selected from a defined set of nucleic acid sequences.

21. The system of claim 20, wherein said defined set of DNA sequences has $[X_N]^N$ combinations, where X_N is sequence of deoxyribonucleotides and the number of deoxyribonucleotides in each sequence is N, N is greater than or equal to three.

22. The system of claim 21, where said deoxyribonucleotides are selected from the group consisting of deoxyriboadenosine, deoxyriboguanosine, deoxyribocytidine, and deoxyribothymidine.

23. The system of claim 21, wherein N is 3-20.

24. The system of claim 23, wherein N is 4-10.

25. The system of claim 24, wherein N is 4 and the number of combinations is 256.

26. The system of claim 18, where the screening sequence/binding protein is selected from the group consisting of EBV origin of replication/EBNA, HSV origin of replicati n/UL9, VZV rigin of replication/UL9-like, and HPV origin of replication/E2, and lambda o_L-o_R/cro.

27. The system of claim 26, wherein the DNA screening sequence is from the HSV origin of replication and the binding protein is UL9.

28. The system of claim 27, wherein the DNA screening sequence is selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:15 and SEQ ID NO:17.

29. The system of claim 28, where the DNA screening sequence is SEQ ID NO:1.

30. The system of claim 29, where the U residue in position 8 is biotinylated.

31. The system of claim 30, where said detection means includes streptavidin, and the streptavidin is bound to a solid support.

32. The system of claim 31, where streptavidin is used to capture the duplex DNA when it is free of bound protein.

33. A method for inhibiting the binding of a DNA-binding protein to duplex DNA, comprising

contacting a compound with a duplex DNA which contains a test sequence adjacent a screening sequence, where the DNA binding protein is effective to bind to the screening sequence with a binding affinity that is substantially independent of said test sequence, further where the binding of said compound to the test sequence inhibits the binding of the protein to the screening sequence.

34. The method of claim 33, wherein the comp und is identified by the steps of

preparing a series of duplex nucleic acid fragments,

each containing a test sequence composed of one of the 4N possible permutations of sequences in a sequence of bas pairs having N-basepairs, where said test sequence is adjacent the screening sequence,

measuring the binding affinity of the DNA binding protein to each of the series of nucleic acid fragments in the presence of the compound, and

selecting the compound if it lowers the binding affinity of the DNA binding protein for the screening sequence.

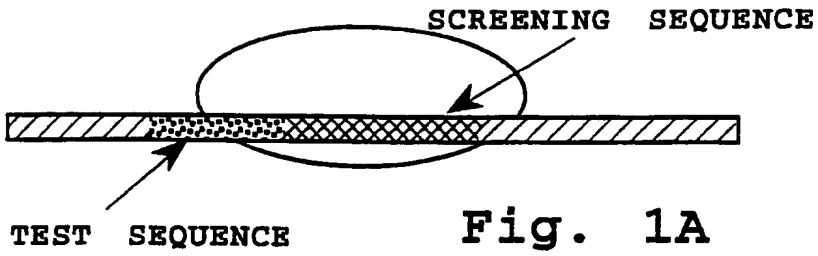


Fig. 1A

Fig. 1B

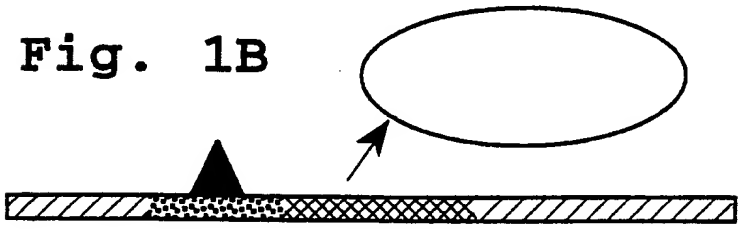
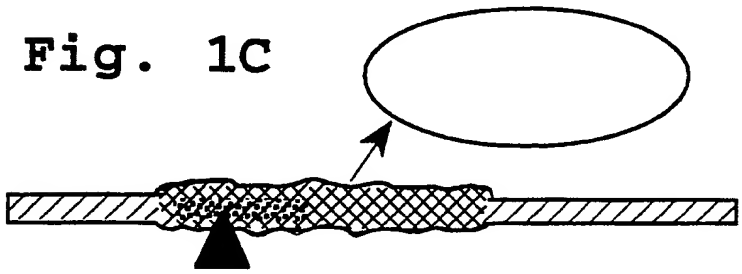


Fig. 1C



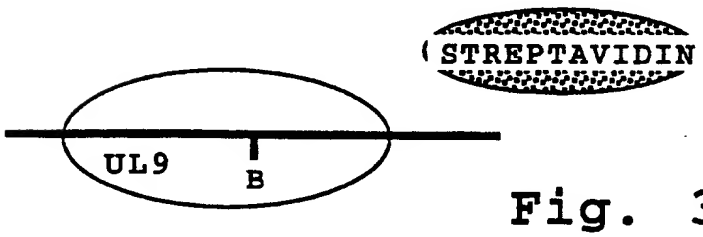


Fig. 3

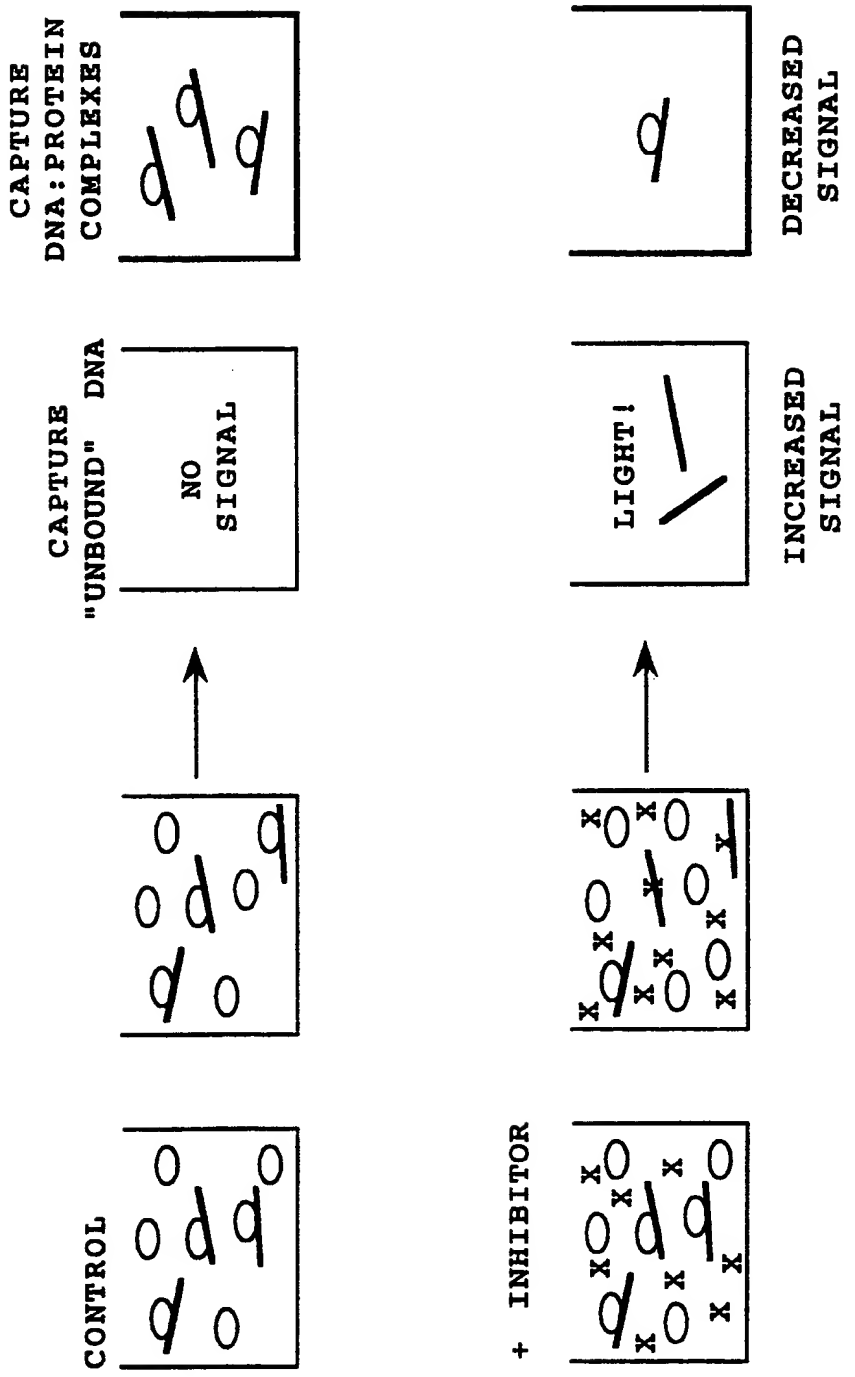

CAPTURE DNA:PROTEIN COMPLEXES
CAPTURE "UNBOUND" DNA
NO SIGNAL
LIGHT!
INCREASED SIGNAL
DECREASED SIGNAL
CONTROL
+ INHIBITOR

Fig. 2

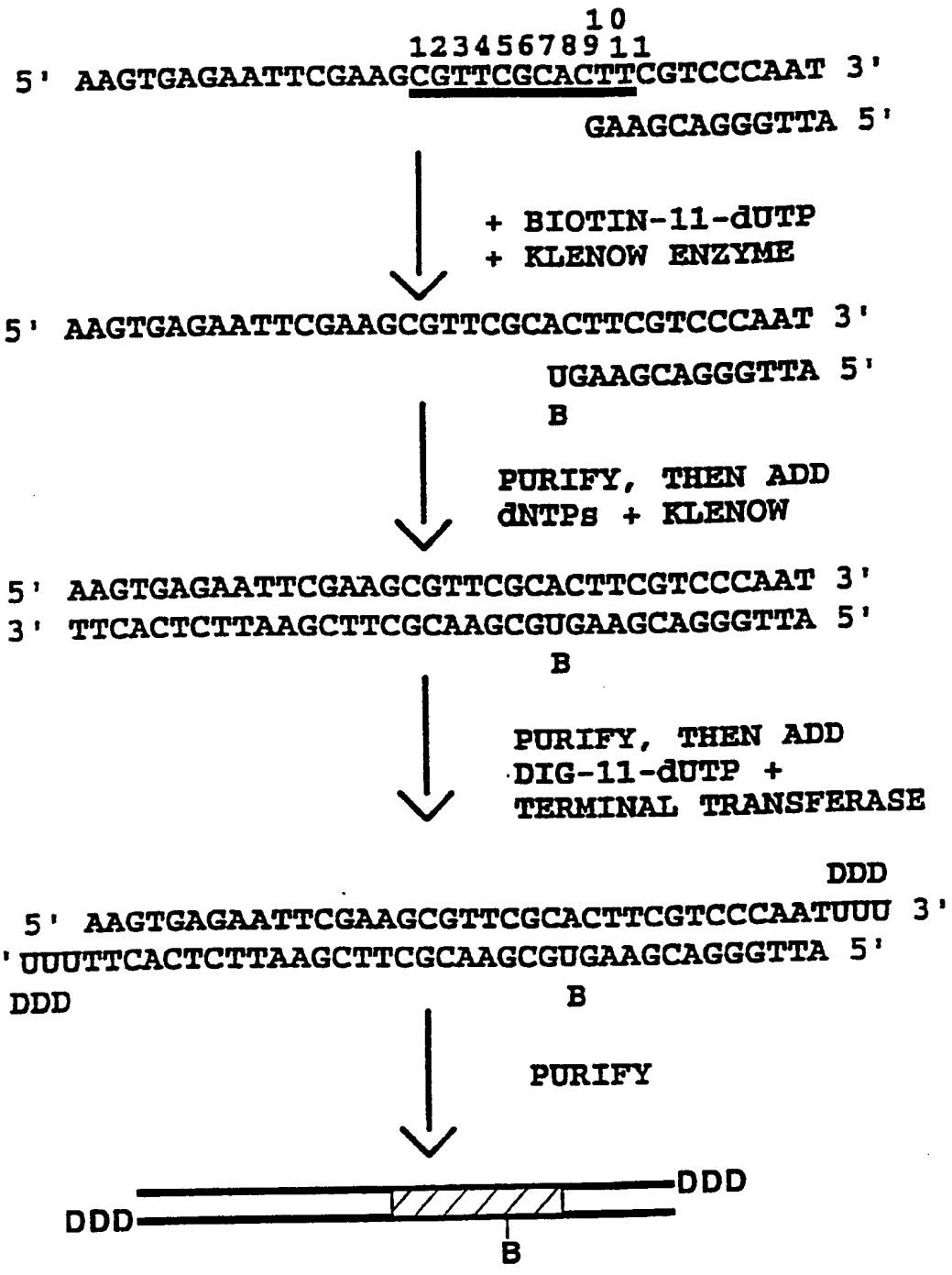

10
1234567891 1
5' AAGTGAGAATTCGAAGCGTTCGCACTTCGTCCCAAT 3'
GAAGCAGGGTTA 5'
+ BIOTIN-11-dUTP
+ KLENOW ENZYME
5' AAGTGAGAATTCGAAGCGTTCGCACTTCGTCCCAAT 3'
UGAAGCAGGGTTA 5'
B
PURIFY, THEN ADD
dNTPs + KLENOW
5' AAGTGAGAATTCGAAGCGTTCGCACTTCGTCCCAAT 3'
3' TTCACTCTTAAGCTTCGCAAGCGUGAAGCAGGGTTA 5'
B
PURIFY, THEN ADD
DIG-11-dUTP +
TERMINAL TRANSFERASE
DDD
5' AAGTGAGAATTCGAAGCGTTCGCACTTCGTCCCAATUUU 3'
3' UUUTTCACTCTTAAGCTTCGCAAGCGUGAAGCAGGGTTA 5'
DDD
B
PURIFY
DDD
DDD
B

Fig. 4

	Test Sequence:	Screening Sequence: Test Sequence
UL9Z1	5′-GCGCGCGCGCGTTCGCACTTCCGCCGCCGG-3′	Z-DNA
UL9Z2	5′-GGCGCCGGCCGTTCGCACTTCGCGCGCGCG-3′	Z-DNA
UL9 CCCG	5′-GGCCCGCCCCGTTCGCACTTCCCGCCCCGG-3′	
UL9 GGGC	5′-GGCGGGCGCCGTTCGCACTTGGGCGGGCGG-3′	
UL9 ATAT	5′-GGATATATACGTTCGCACTTTAATTATTGG-3′	
UL9 polyA	5′-GGAAAAAAACGTTCGCACTTAAAAAAAAGG-3′	
UL9 polyT	5′-GGTTTTTTTCGTTCGCACTTTTTTTTTTGG-3′	
UL9 GCAC	5′-GGACGCACGCGTTCGCACTTGCAGCAGCGG-3′	
ATori-1	5′-GCGTATATATCGTTCGCACTTCGTCCCAAT-3′	
oriEco2	5′-GGCGAATTCGACGTTCGCACTTCGTCCCAAT-3′	
oriEco3	5′-GGCGAATTCGATCGTTCGCACTTCGTCCCAAT-3′	

Fig. 5

5/16

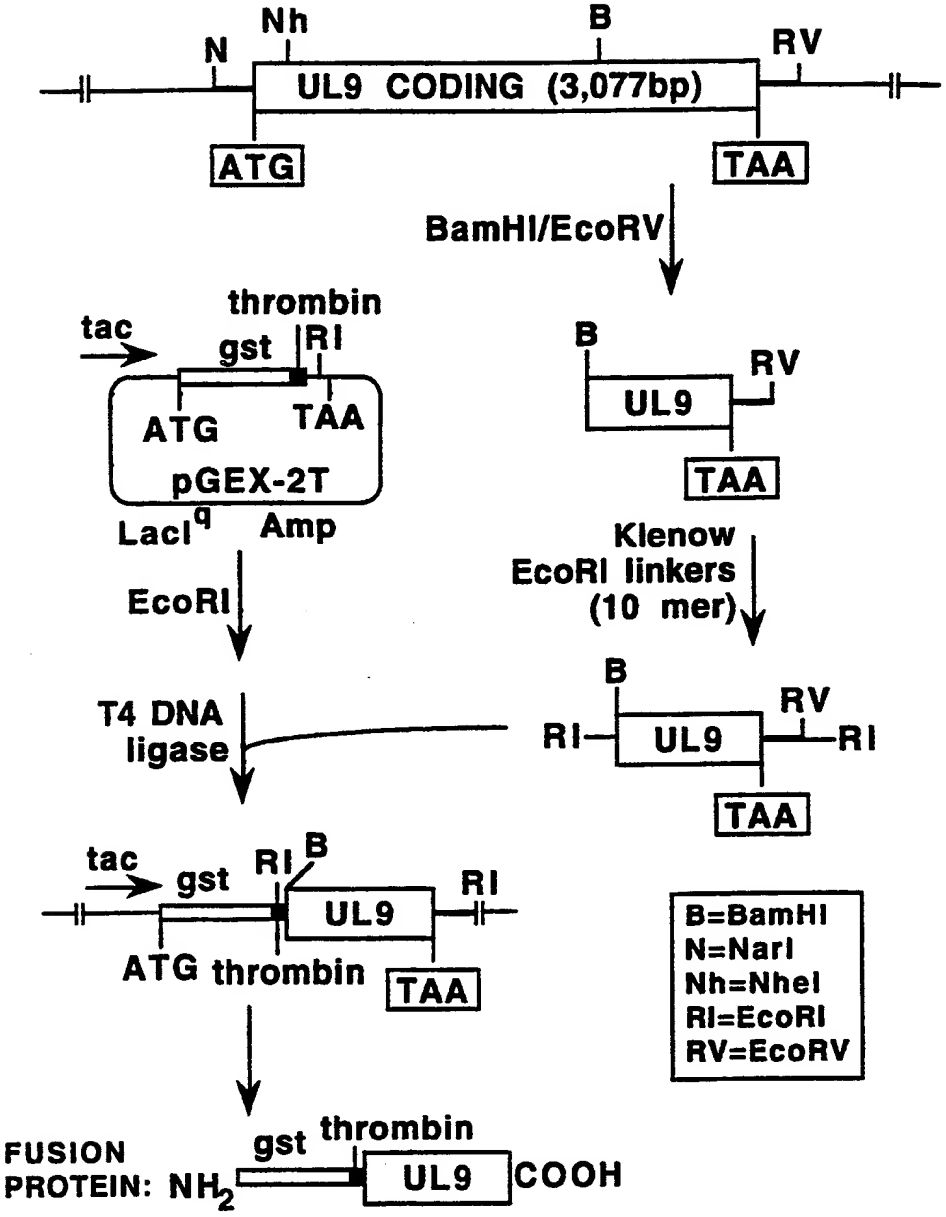

Nh
B
N
RV
UL9 CODING (3,077bp)
ATG
TAA
BamHI/EcoRV
thrombin
tac
gst
RI
ATG
TAA
pGEX-2T
LacI^q
Amp
B
RV
UL9
TAA
Klenow
EcoRI linkers
(10 mer)
EcoRI
T4 DNA
ligase
B
RV
RI
UL9
RI
TAA
tac
gst
RI
B
UL9
RI
ATG thrombin
TAA
B=BamHI
N=NarI
Nh=NheI
RI=EcoRI
RV=EcoRV
FUSION
PROTEIN: NH2
gst
thrombin
UL9
COOH
28 kdal + 35 kdal = 63 kdal
(255 aa) (316 aa)

Fig. 6

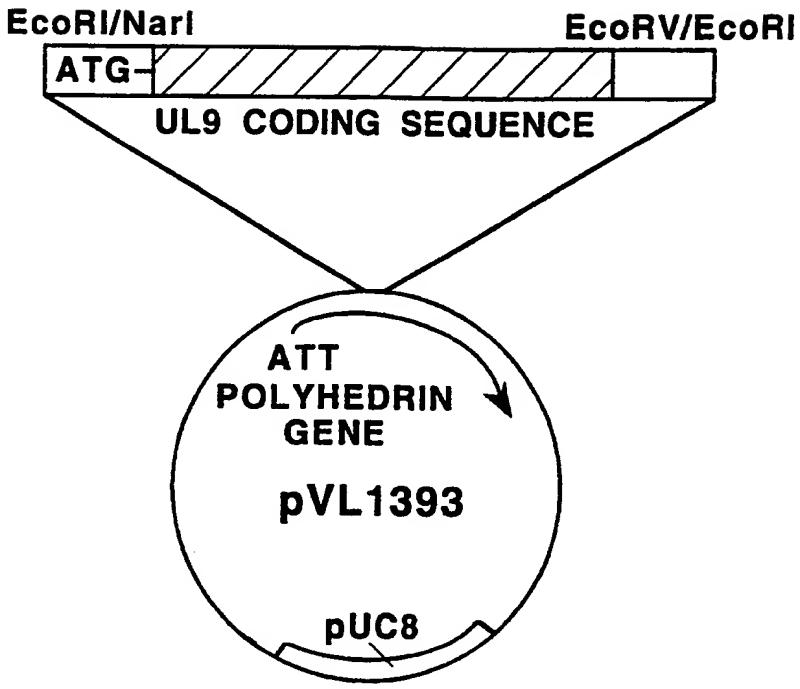

EcoRI/NarI
EcoRV/EcoRI
ATG
UL9 CODING SEQUENCE
ATT
POLYHEDRIN
GENE
pVL1393
pUC8

Fig. 7

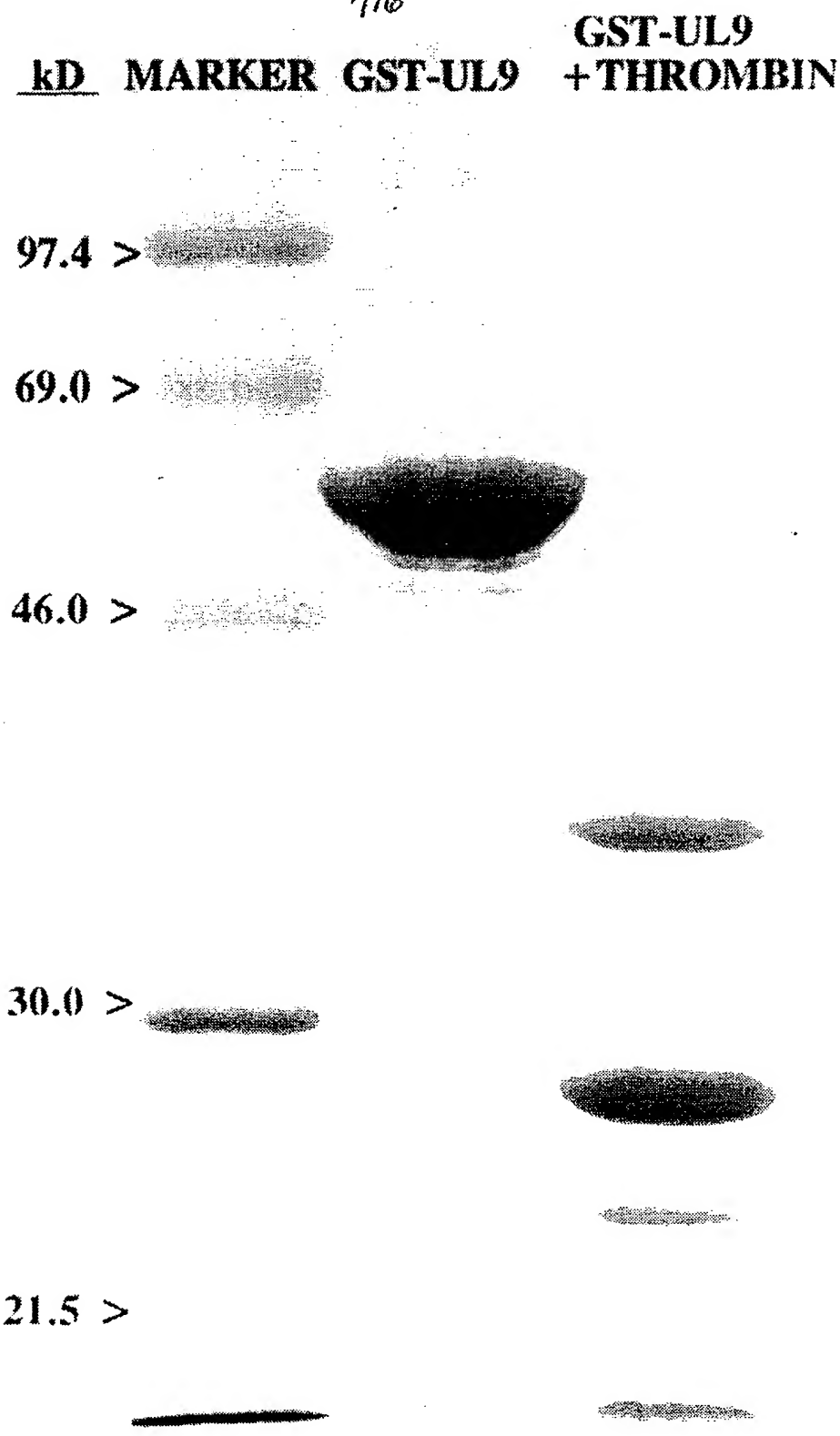


Fig. 8

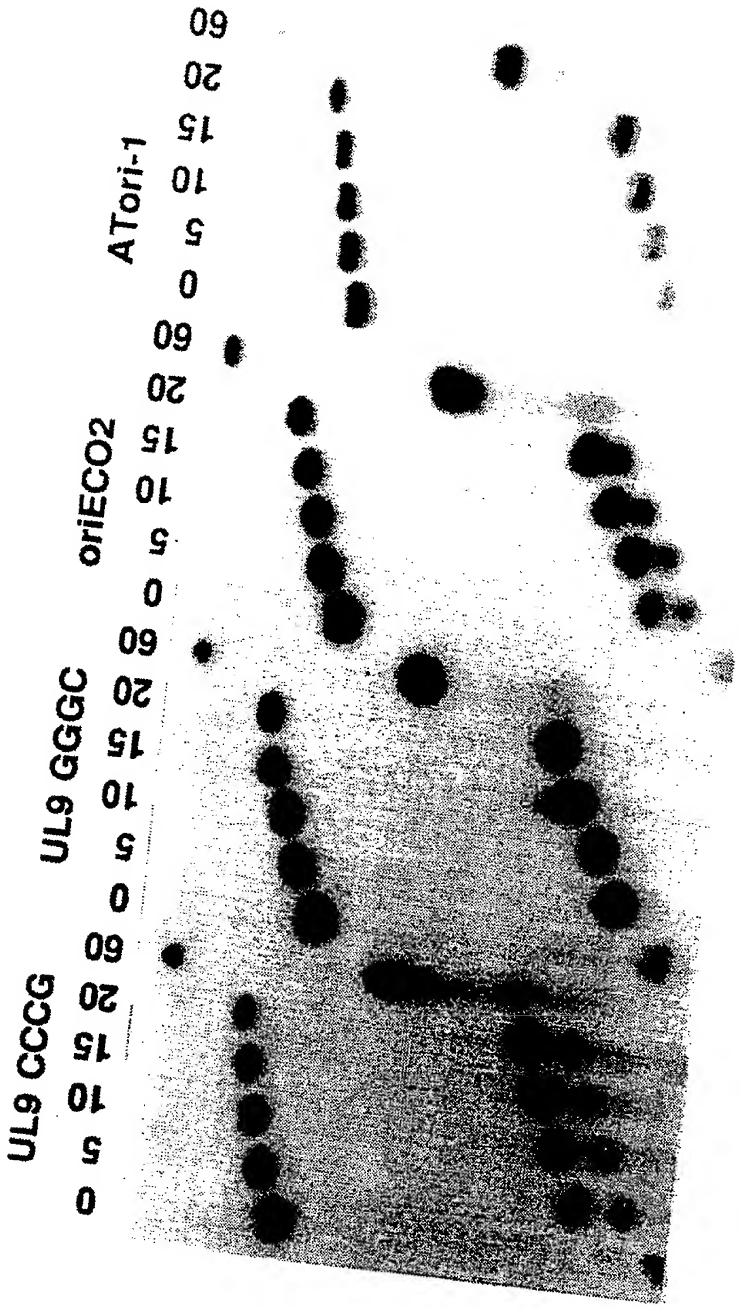


Fig. 9

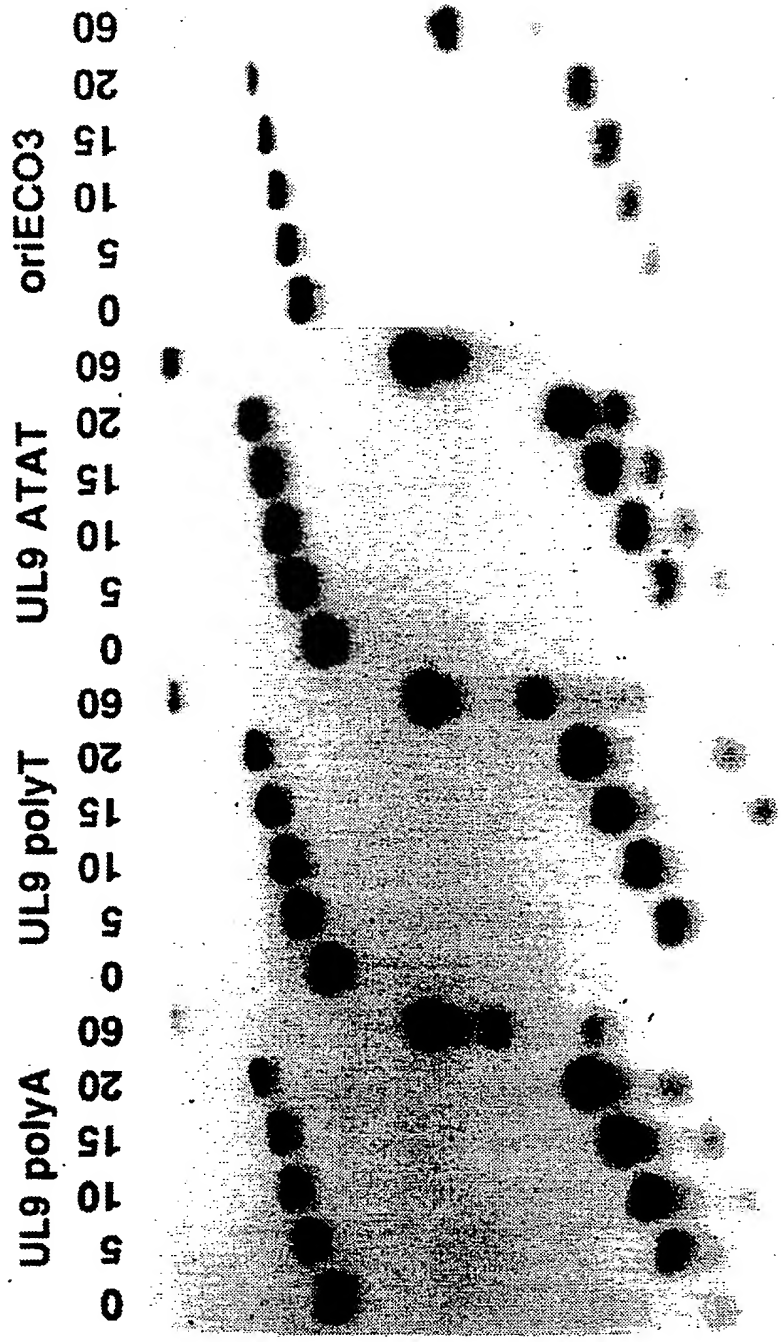
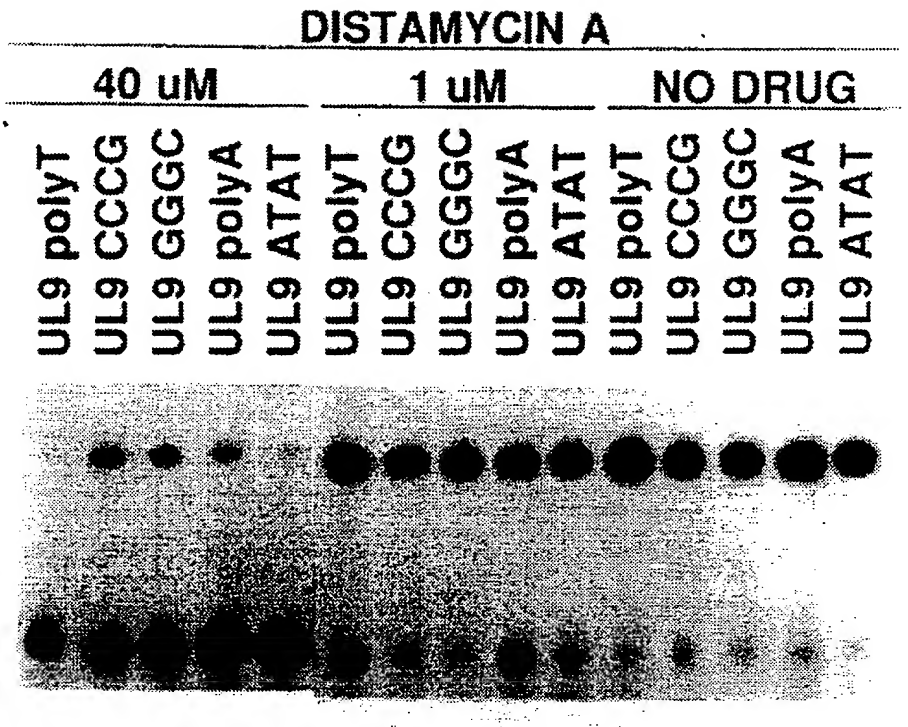


Fig. 9 (cont)


DISTAMYCIN A
40 uM
1 uM
NO DRUG
UL9 polyT
UL9 CCCG
UL9 GGGC
UL9 polyA
UL9 ATAT
UL9 polyT
UL9 CCCG
UL9 GGGC
UL9 polyA
UL9 ATAT
UL9 polyT
UL9 CCCG
UL9 GGGC
UL9 polyA
UL9 ATAT

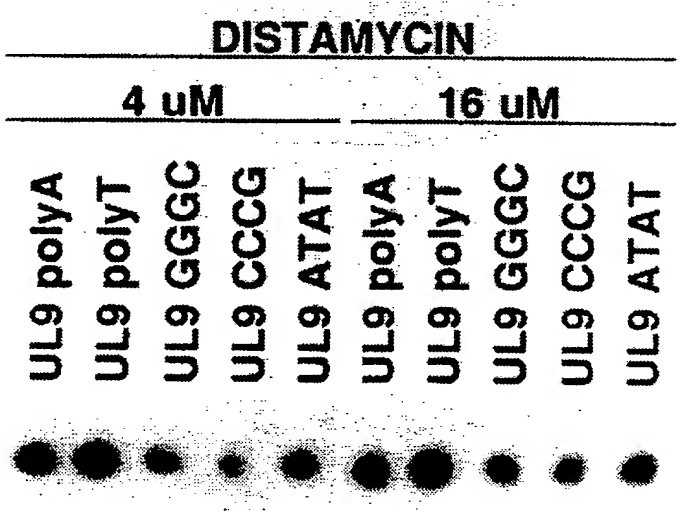

DISTAMYCIN
4 uM
16 uM
UL9 polyA
UL9 polyT
UL9 GGGC
UL9 CCCG
UL9 ATAT
UL9 polyA
UL9 polyT
UL9 GGGC
UL9 CCCG
UL9 ATAT

Fig. 10A

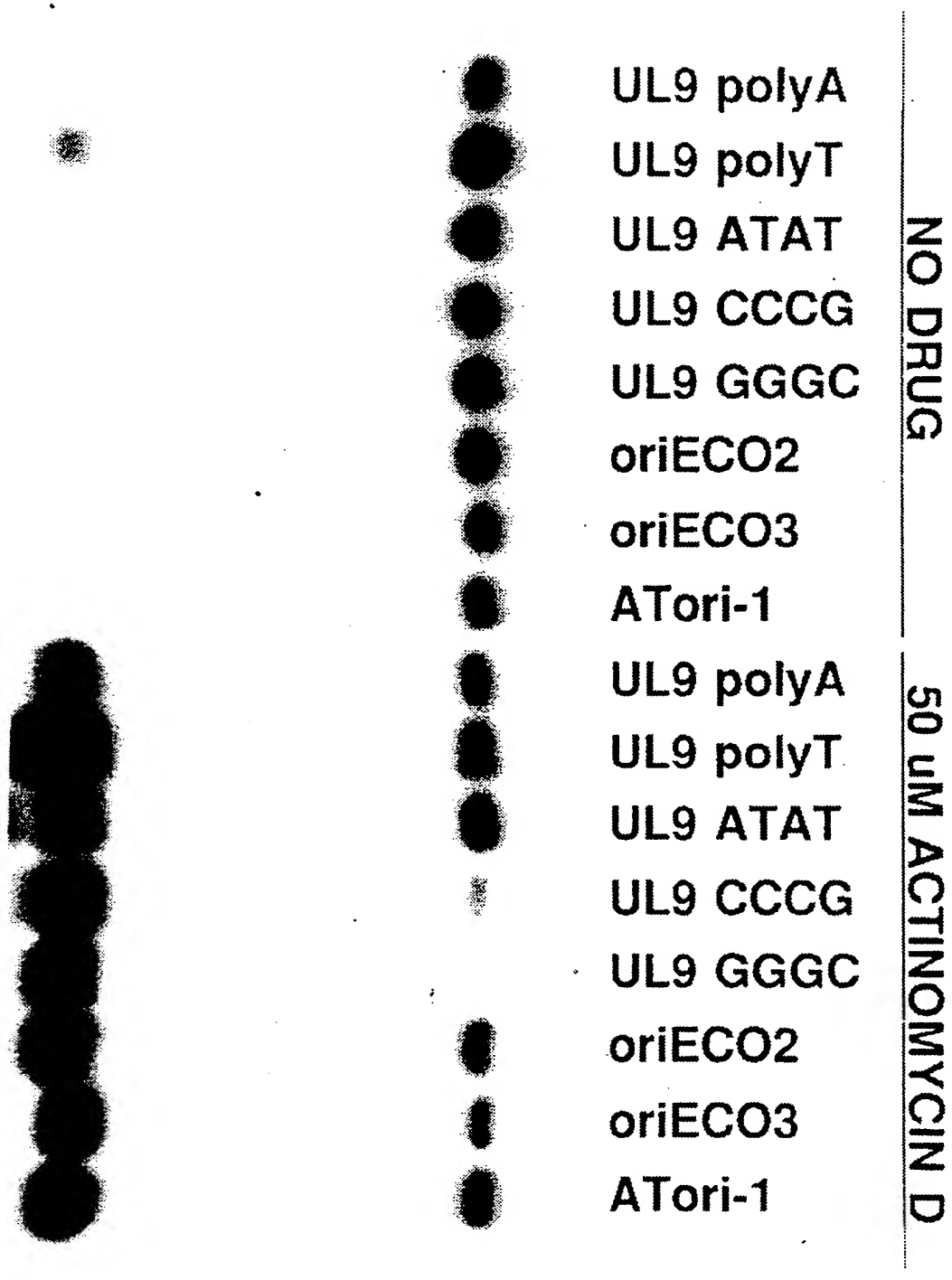

UL9 polyA
UL9 polyT
UL9 ATAT
UL9 CCCG
UL9 GGGC
oriECO2
oriECO3
ATori-1
NO DRUG
UL9 polyA
UL9 polyT
UL9 ATAT
UL9 CCCG
UL9 GGGC
oriECO2
oriECO3
ATori-1
50 uM ACTINOMYCIN D

Fig. 10B

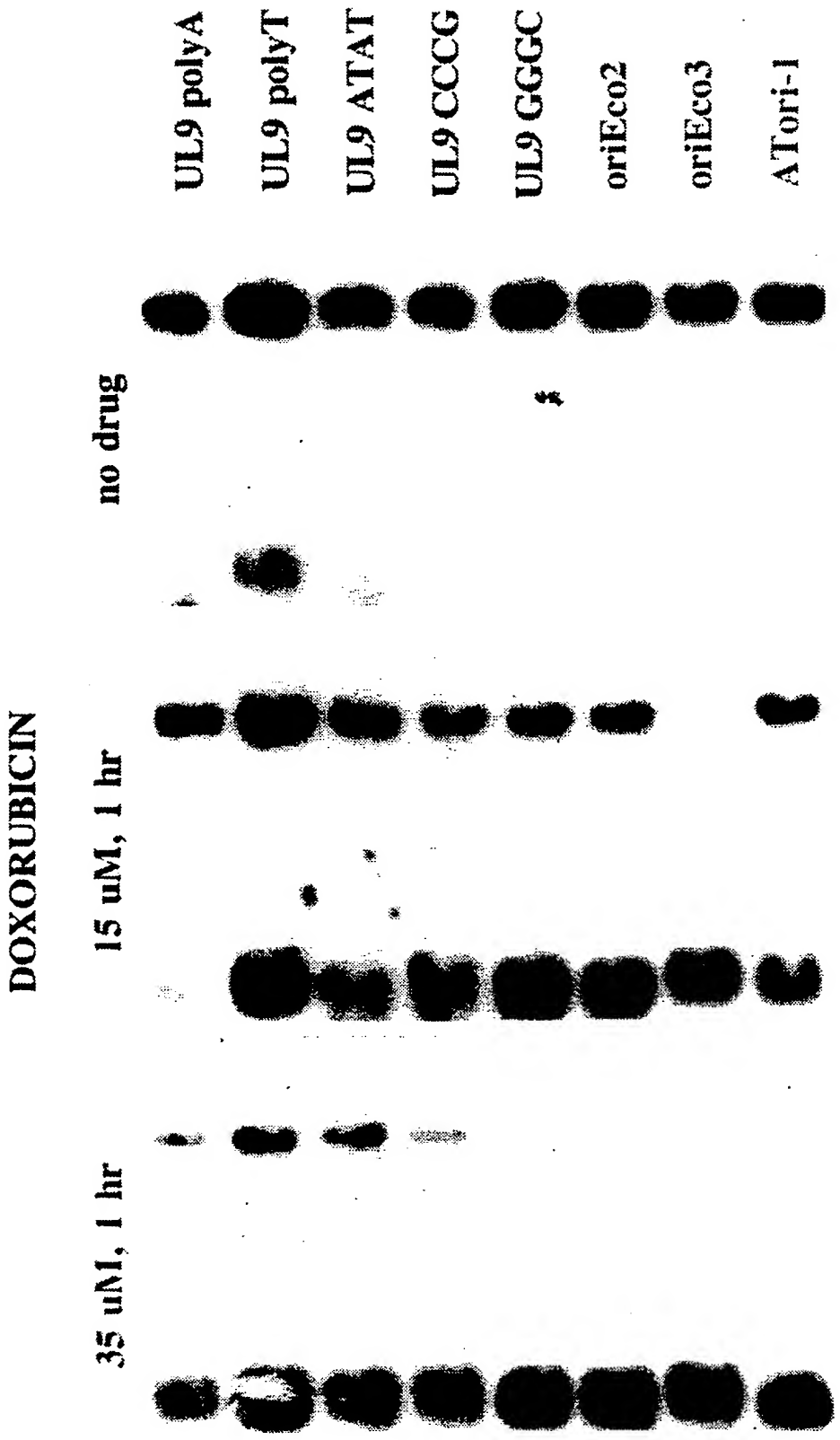

UL9 polyA
UL9 polyT
UL9 ATAT
UL9 CCCG
UL9 GGGC
oriEco2
oriEco3
ATori-1
DOXORUBICIN
no drug
15 uM, 1 hr
35 uM, 1 hr

Fig. 10C

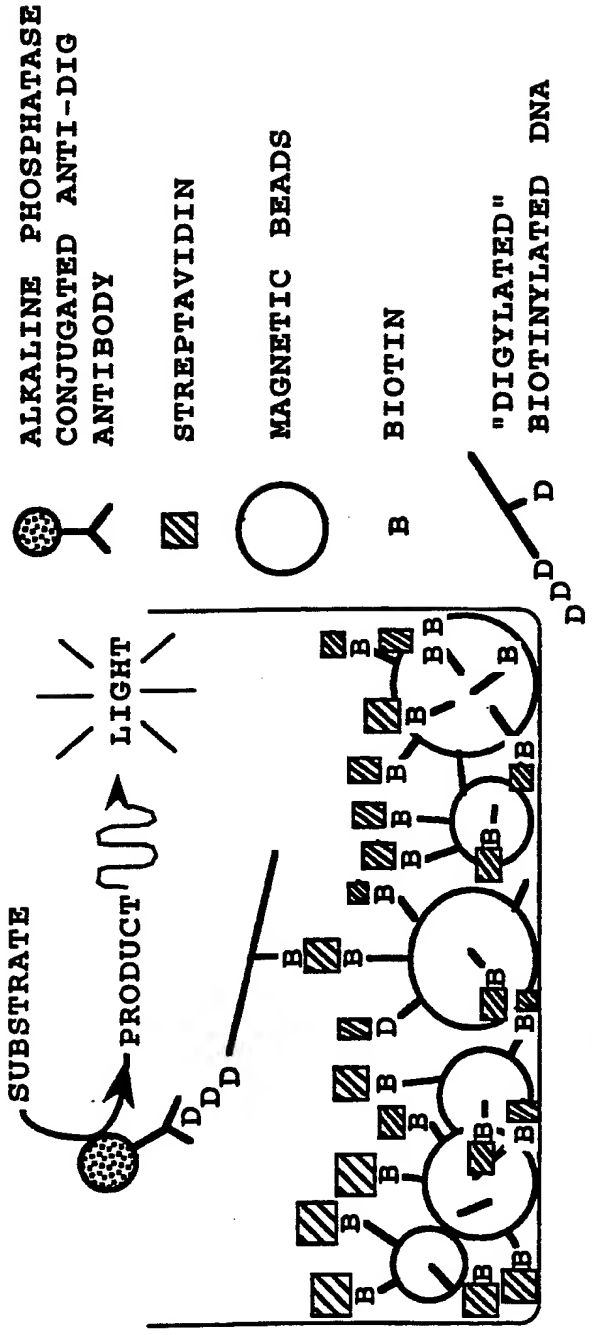


FIG. 11A

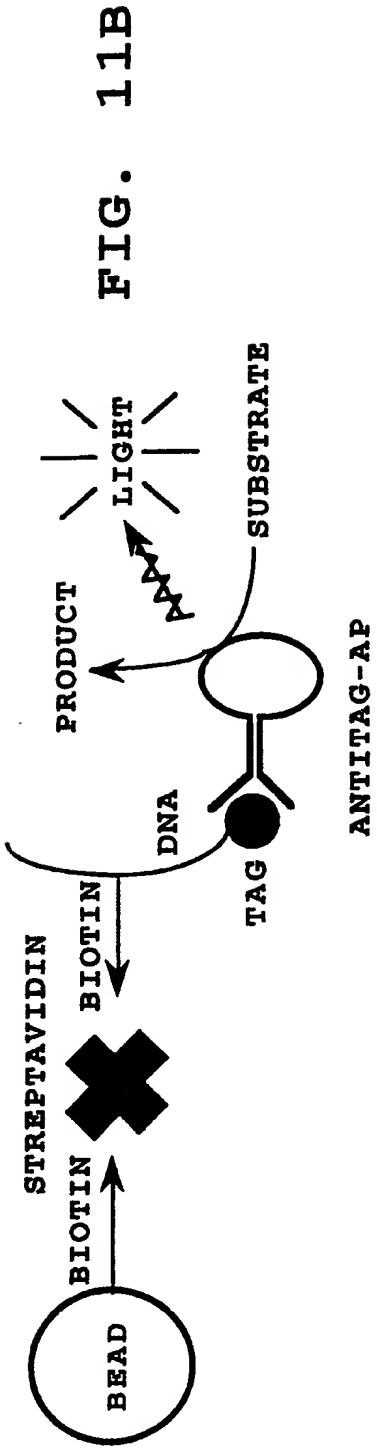


FIG. 11B

Sequence		Test Mixtures 1	2	3	4	5	6	7	8
AGCTTTCGCACTTAGCT	-	+	+	-	-	-	-	-	-
AGCATTCGCACTTAGCA	-	+	-	-	+	+	-	-	-
AGCCTTCGCACTTAGCC	-	+	-	-	-	-	-	-	+
AGCGTTCGCACTTAGCG	-	+	-	-	-	-	-	-	-
TGCTTTCGCACTTTGCT	-	+	-	-	-	-	-	+	-
TGCATTCGCACTTTGCA	-	+	-	-	+	+	-	+	-
TGCCTTCGCACTTTGCC	-	+	-	-	-	-	-	+	+
TGCGTTCGCACTTTGCG	-	+	-	-	-	-	-	+	-
:	:	:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:		
CCATTTCGCACTTCCAT	-	+	+	-	+	-	-	-	+
CCCTTTCGCACTTCCCT	-	+	+	-	-	-	-	+	+
CCGTTTCGCACTTCCGT	-	+	+	-	-	-	-	-	+
CCTTTTCGCACTTCCTT	-	+	+	-	-	-	-	+	+
:	:	:	:	:	:	:	:	:	:

Fig. 12

SEQUENCE				
AAAA 001	ATCA 053	CGGA 105	GCTA 157	TCAA 209
AAAC 002	ATCC 054	CGCG 106	GCTC 158	TCAC 210
AAAG 003	ATCG 055	CGGG 107	GCTG 159	TCAG 211
AAAT 004	ATCT 056	CGGT 108	GCTT 160	TCAT 212
AACA 005	ATGA 057	CGTA 109	GGAA 161	TCCA 213
AACC 006	ATGC 058	CGTC 110	GGAC 162	TCCC 214
AACG 007	ATGG 059	CGTG 111	GGAG 163	TCCG 215
AACT 008	ATGT 060	CGTT 112	GGAT 164	TCCT 216
AAGA 009	ATTA 061	CTAA 113	GGCA 165	TCGA 217
AAGC 010	ATTC 062	CTAC 114	GGCC 166	TCGC 218
AAGC 011	ATTG 063	CTAG 115	GGCG 167	TCGG 219
AAGT 012	ATTT 064	CTAT 116	GGCT 168	TCGT 220
AATA 013	CAAA 065	CTCA 117	GGGA 169	TCTA 221
AATC 014	CAAC 066	CTCC 118	GGGC 170	TCTC 222
AATG 015	CAAG 067	CTCG 119	GGGG 171	TCTG 223
AATT 016	CAAT 068	CTCT 120	GGGT 172	TCTT 224
ACAA 017	CACA 069	CTGA 121	GGTA 173	TGAA 225
ACAC 018	CACC 070	CTGC 122	GGTC 174	TGAC 226
ACAG 019	CACG 071	CTGG 123	GGTG 175	TGAG 227
ACAT 020	CACT 072	CTGT 124	GGTT 176	TGAT 228
ACCA 021	CAGA 073	CTTA 125	GTAA 177	TGCA 229
ACCC 022	CAGC 074	CTTC 126	GTAC 178	TGCC 230
ACCG 023	CAGG 075	CTTG 127	GTAG 179	TGCG 231
ACCT 024	CAGT 076	CTTT 128	GTAT 180	TGCT 232
ACGA 025	CATA 077	GAAA 129	GTCA 181	TGGA 233
ACGC 026	CATC 078	GAAC 130	GTCC 182	TGGC 234
ACGC 027	CATG 079	GAAG 131	GTCG 183	TGGG 235
ACGT 028	CATT 080	GAAT 132	GTCT 184	TGGT 236
ACTA 029	CCAA 081	GACA 133	GTGA 185	TGTA 237
ACTC 030	CCAC 082	GACC 134	GTGC 186	TGTC 238
ACTG 031	CCAG 083	GACG 135	GTGG 187	TGTG 239

Fig 13

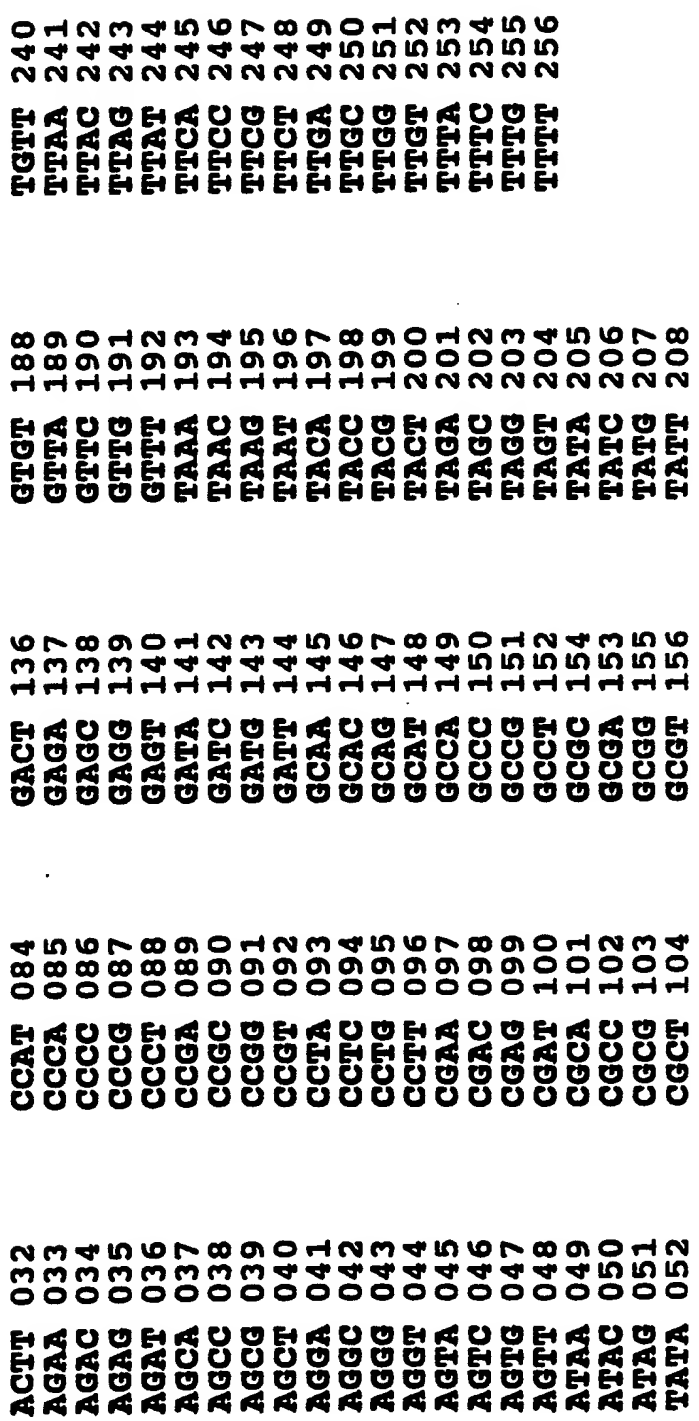

ACTT 032 CCAT 084 GACT 136 GTGT 188 TGTT 240
AGAA 033 CCCA 085 GAGA 137 GTTA 189 TTAA 241
AGAC 034 CCCC 086 GAGC 138 GTTC 190 TTAC 242
AGAG 035 CCCG 087 GAGG 139 GTTG 191 TTAG 243
AGAT 036 CCCT 088 GAGT 140 GTTT 192 TTAT 244
AGCA 037 CCGA 089 GATA 141 TAAA 193 TTCA 245
AGCC 038 CCGC 090 GATC 142 TAAC 194 TTCC 246
AGCG 039 CCGG 091 GATG 143 TAAG 195 TTCG 247
AGCT 040 CCGT 092 GATT 144 TAAT 196 TTCT 248
AGGA 041 CCTA 093 GCAA 145 TACA 197 TTGA 249
AGGC 042 CCTC 094 GCAC 146 TACC 198 TTGC 250
AGGG 043 CCTG 095 GCAG 147 TACG 199 TTGG 251
AGGT 044 CCTT 096 GCAT 148 TACT 200 TTGT 252
AGTA 045 CGAA 097 GCCA 149 TAGA 201 TTTA 253
AGTC 046 CGAC 098 GCCC 150 TAGC 202 TTTC 254
AGTG 047 CGAG 099 GCCG 151 TAGG 203 TTTG 255
AGTT 048 CGAT 100 GCCT 152 TAGT 204 TTTT 256
ATAA 049 CGCA 101 GCGC 154 TATA 205
ATAC 050 CGCC 102 GCGA 153 TATC 206
ATAG 051 CGCG 103 GCGG 155 TATG 207
TATA 052 CGCT 104 GCGT 156 TATT 208

Fig. 13 (con't)

GATC	AGTC	TAGC	CGAT
GACT	AGCT	TACG	CGTA
GTCA	ATCG	TGCA	CATG
GTAC	ATGC	TGAC	CAGT
GCTA	ACTG	TCAG	CTAG
GCAT	ACGT	TCGA	CTGA

Fig. 14

TEST SCREEN TEST

5'-GCGTAAXXXXCGTTCGCACTTXXXXCTTCGTCCCAAT-3'

3'-CGCATTYYYYGCAAGCGTGAAYYYYGAAGCAGGGTTA-5'

PRIMER

Fig. 15

INTERNATIONAL SEARCH REPORT

International Application No PCT/US 92/05476

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all)[6]

According to International Patent Classification (IPC) or to both National Classification and IPC

Int.Cl. 5 C12Q1/68

II. FIELDS SEARCHED

Minimum Documentation Searched[7]

Classification System	Classification Symbols
Int.Cl. 5	C12Q ; G01N ; C12N

Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched[8]

III. DOCUMENTS CONSIDERED TO BE RELEVANT[9]

Category °	Citation of Document, [11] with indication, where appropriate, of the relevant passages [12]	Relevant to Claim No.[13]
A	WO,A,8 704 170 (MASSACHUSSETTS INSTITUTE OF TECHNOLOGY) 16 July 1987 see page 11, line 1 - page 13, line 30 see page 38, line 15 - page 40, line 7 ---	1
A	US,A,4 257 774 (C.L.RICHARDSON ET AL.) 24 March 1981 see the whole document ---	1
A	US,A,4 270 924 (S.T.CROOKE ET AL.) 2 June 1981 see column 1, line 50 - line 67 see column 3, line 15 - column 4, line 2 ---	1

-/--

° Special categories of cited documents :[10]

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search	Date of Mailing of this International Search Report
13 OCTOBER 1992	21. 10. 92
International Searching Authority	**Signature of Authorized Officer**
EUROPEAN PATENT OFFICE	LUZZATTO E.R.

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)

Category °	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claim No.
A	ANALYTICAL BIOCHEMISTRY vol. 193, no. 2, March 1991, NEW YORK US pages 220 - 224 K.HOBSON ET AL. 'Use of DNA-proteion interaction to isolate specific genomic DNA sequences' see abstract see "Discussion" ---	1,5
A	NUCLEIC ACIDS RESEARCH. vol. 18, no. 1, 11 January 1990, ARLINGTON, VIRGINIA US pages 157 - 161 J.C.HANVEY ET AL. 'Site-specific inhibition of EcoRI restriction/modification enzymes by a DNA triple helix' see abstract see page 161, left column, line 1 - line 25 -----	1,18

ANNEX TO THE INTERNATIONAL SEARCH REPORT ON INTERNATIONAL PATENT APPLICATION NO. US 9205476 SA 62049

This annex lists the patent family members relating to the patent documents cited in the above-mentioned international search report. The members are as contained in the European Patent Office EDP file on 13/10/92
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
WO-A-8704170	16-07-87	EP-A-	0252146	13-01-88
		JP-T-	63502480	22-09-88
US-A-4257774	24-03-81	None		
US-A-4270924	02-06-81	CA-A-	1129321	10-08-82
		AT-B-	365343	11-01-82
		AU-B-	536522	10-05-84
		AU-A-	5719380	16-10-80
		BE-A-	882669	06-10-80
		CH-A-	644453	31-07-84
		DE-A-	3013837	30-10-80
		FR-A,B	2454099	07-11-80
		GB-A,B	2047712	03-12-80
		JP-A-	55143441	08-11-80
		LU-A-	82345	16-12-80
		NL-A-	8002094	14-10-80
		SE-A-	8002602	09-12-80